Psychological Bulletin

Harry Helson, Editor
University of Texas

CONTENTS

This is the last issue of Volume 57
Volume contents and title page appear herein.

Published Bimonthly by the
American Psychological Association

Vol. 57, No. 6 November 1960

The *Psychological Bulletin* contains evaluative reviews of research literature and reviews of research methodology and instrumentation in psychology. This Journal *does not publish reports of original research or original theoretical articles.*

Manuscripts should be sent to the Editor, Harry Helson, Department of Psychology, University of Texas, Austin 12, Texas.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1957 Revision). Special attention should be given to the section on the preparation of the references (pp. 50–60), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted *in duplicate,* one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

Reprints. Fifty free reprints are given to contributors of articles and notes.

Communications—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: \$8.00 (Foreign \$8.50). Single copies, \$1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Menasha, Wisconsin
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second class matter at Washington, D.C., and Menasha, Wisconsin. Second-class postage paid at Washington, D.C., and at additional mailing offices. Printed in U.S.A.

VOL. 57, NO. 6 NOVEMBER 1960

Psychological Bulletin

RETROACTIVE AND PROACTIVE INHIBITION OF VERBAL LEARNING[1]

NORMAN J. SLAMECKA *University of Vermont* AND JOHN CERASO *Yeshiva University*

The last review of the literature solely devoted to retroactive inhibition (RI) was Swenson's (1941) monograph whose coverage extended through 1940. The present paper extends the coverage by presenting a full bibliography and critical analysis of all published reports on the RI and proactive inhibition (PI) of verbal learning from 1941 through 1959. Studies of infrahuman *S*s and of nonverbal behavior were excluded because of considerations of length and the fact that, traditionally, RI is a concept associated with verbal behavior. Excluded also were studies using interpolated convulsive seizures or surgical procedures because such treatments are qualitatively different from intervening learning as such and require other theoretical formulations to explain their effects. Following a brief summary of the field in 1940, subsequent developments will be discussed under five general headings: Degree of Acquisition, Similarity of Materials, Extrinsic Factors, Temporal Effects, Major Theoretical Positions.

The dominant theoretical position in 1940 was a transfer theory, given its fullest exposition by McGeoch and his collaborators. In essence the theory stated that RI could be explained by the general principles discovered in the study of transfer. The failure of performance of an old association could be attributed to greater strength of the new association, a mutual blocking of old and new associations, or a confusion between the two.

This theory was capable of handling a great deal of the relevant data and depended largely upon two sources of evidence for empirical support. The first source was the evidence for the effect of similarity of materials upon RI, which supported the contention that RI could be explained by the principles of transfer. The second source was intrusion errors, which are responses from the interpolated learning offered by *S*s when they are asked for responses from the original learning. The existence of these errors supported the contention that old responses were not given because new ones had supplanted them.

Much of the subsequent history of RI can be viewed as a process of extension and enlargement of McGeoch's basic position. The four major theories discussed later on in this paper serve as leading examples. The Melton-Irwin two-factor theory enlarged the competition of response theory by postulating an unlearning process in addition to competition of response. Gibson elaborated the theory by placing it within the setting of the

[1] This work was supported in part by a grant to the senior author from the National Science Foundation (G-6192).

conditioning experiment, making available the conceptual apparatus of differentiation and generalization. Underwood's work has concentrated upon clarifying the nature of both unlearning and differentiation, while Osgood has stressed the communality of transfer and RI in his "transfer and retroaction surface."

A consideration of terms is now in order. RI is the decrement in retention attributable to interpolated learning (McGeoch & Irion, 1952), and the operations that define it require a comparison of the retention of some original learning (OL) between two groups that differ in some aspect of the interpolated activity (IL) (Underwood, 1949a). The experimental group has IL, and the control group engages in some nonlearning filler task. Better retention in the control group defines RI, and better retention in the experimental group defines retroactive facilitation. Since the control group almost always shows some loss of the OL after its "rest activity," to what can the decrement be attributed: to incidental learning, to loss of set, to sheer metabolic activity (Shaklee & Jones, 1959)? The impossibility of assuring that no interpolated learning takes place for that group introduces an inevitable looseness into the significance of the RI measure. The control group's decrement is sometimes assumed to be due to "natural" forgetting, as distinct from the additional decrement attributed to the specific interfering tasks given the experimental group. But if a strict interference position is to be maintained, the "natural" forgetting must also be attributed to some source of interference, albeit beyond *E*'s control. The fact that different investigators may employ different filler tasks imposes a shifting base against which experimentally induced RI is calculated and renders comparison of results difficult. Osgood (1946, 1948) has dealt with the problem by simply omitting the control group and regarding RI as the difference in performance between the end of OL and the subsequent OL relearning (RL), lumping together both the specific and nonspecific decremental variables operating during the interpolated interval. This, of course, is a measure of total forgetting. Such a straightforward procedure cannot, however, distinguish between RI and retroactive facilitation, as they are usually understood, since facilitation may involve simply less decrement in retention as compared to a control group. Another troublesome problem arises with the other methods of quantifying RI, both of which rely upon control groups. Absolute RI is simply the numerical difference between the retention of the control and experimental groups, and relative RI is the percentage difference between them:

$$\frac{\text{Rest-Work} \times 100}{\text{Rest}}$$

Each of these measures is thus dually dependent upon both the experimental and the control groups' performance, and they may not always give the same pattern of results. This problem becomes especially important in studies of degree of OL upon RI. It is often the case that as OL increases, absolute RI increases, but relative RI decreases (Postman & Riley, 1959). To illustrate, it can be seen that, when degree of OL is low, the control group's retention is low, and even slight departures from this baseline on the part of the experimental group will represent a substantial percentage difference; where-

as when the control's recall is high, the same absolute difference will reflect a lesser percentage change, and the relative RI will have decreased, while absolute RI will have remained the same. At present, we can only be alerted to this source of confusion and take it into account when viewing the results of any RI study. The foregoing observations apply just as fully to the quantification of PI, to which we now turn.

The PI paradigm requires a comparison of the retention of some original learning (List 2) between two groups that differ only in some aspect of the activity preceding that learning. The experimental group learns some previous material (List 1), and the control group does not. The same problem with regard to the control group's experience applies here. Better retention in the control group defines PI, and better retention in the experimental group defines proactive facilitation. In addition, the PI design requires that a clear temporal distinction be made between the end of the acquisition phase of List 2 and its subsequent retention test. Minimally, a retention interval longer than the OL intertrial interval is needed. If this is not done, the learning and retention phases would be operationally identical, and the PI design would be indistinguishable from the transfer design.

Degree of Acquisition

Swenson's (1941) generalizations about the acquisition variables were as follows:

> [*a*] . . . susceptibility to retroaction does not tend to decrease as the amount of original activity is increased . . . (p. 17). [*b*] . . . the greater the degree of learning of the original activity, the less susceptible is the learning to retroactive inhibition (p. 18). [*c*] . . . we may retain the idea of increased retroactive inhibition with increased amount of interpolated activity (p. 19). [*d*] All measures show an increase in retroactive inhibition with early increases in the degree of interpolated learning and a decrease in retroactive inhibition with very high degrees of interpolated learning (p. 20).

These conclusions have been further amplified through subsequent work. (Unless otherwise noted, the results cited below refer to measures at recall—first relearning trial.)

Several papers have reported the effect of degree of IL upon RI either by varying the number of IL trials (Briggs, 1957; Highland, 1949; Melton, 1941; Postman & Riley, 1959; Slamecka, 1959, 1960a; Thune & Underwood, 1943; Underwood 1945, 1950b), by setting a performance criterion (Archer & Underwood, 1951; Osgood, 1948; Richardson, 1956), by varying the number of interfering lists (Underwood, 1945), or by analysis of the associative strength of any single IL list item (Runquist, 1957).

Most of the papers agreed that RI of recall showed a negatively accelerated increase with increasing IL, and studies that carried IL to very high degrees also agreed that the curve tended to flatten out or even to decrease (Briggs, 1957; Thune & Underwood, 1943; Underwood, 1945). In general, maximum levels of RI were obtained when the IL practice had somewhat exceeded the OL practice and further IL trials did not serve to increase the RI appreciably. An exception to this was Runquist's (1957) finding that RI of individual items was not a function of the strength of the corresponding interpolated items. Also, in Exp. B of Underwood's (1945) report, there were no significant recall differences among the work groups, nor was there any consistent trend toward a negatively accelerated curve of recall as a func-

tion of degree of IL. A possible explanation for this may lie in the fact that the lowest IL degree (8 trials) exceeded the mean OL trials (which averaged about 6). Under these conditions it might well be expected that increasing the IL practice would have no further decremental effect. Increasing the IL levels did, however, produce faster RI dissipation, which gives marginal support to Underwood's differentiation hypothesis. The question of whether degree of IL, measured by trials, or amount of IL, measured by the number of different interpolated lists given, is the more powerful variable in producing RI was also specifically tested by Underwood (1945). Care was taken to equate the amount and degree levels by equal total trials, and the findings showed that RI changed at a faster rate with increases in amount than with increases in degree of IL. Both relative and absolute RI grew steadily as the number of IL lists was increased, but the frequency of overt interlist intrusions remained relatively constant, regardless of the number of lists. This is also consistent with the differentiation hypothesis, since increasing the number of lists should not increase differentiation, whereas increasing the number of trials on a single list should increase it. It is urged that a further comparison of the effect of amount against degree of IL should be made, using yet lower IL levels, so as to fill out that part of the curve at which acquisition is very slight.

Degree of OL was controlled in the following studies by varying the number of trials (Briggs, 1957; Melton, 1941; Postman & Riley, 1959; Shaw, 1942; Slamecka, 1960a), setting a performance criterion (Richardson, 1956), or analyzing individual item strengths (Runquist, 1957). All reports agreed that the susceptibility of the original material to RI was inversely related to its level of acquisition. The well-designed factorial study by Briggs (1957), using four OL and five IL levels (2, 5, 10, and 20 trials OL, compared to 0, 2, 5, 10, and 20 trials IL, all paired adjectives), confirmed previous findings as well as showing that, as OL increases, the greater must the IL level be for maximal relative RI. This was also found by Melton (1941). Further, Briggs reported more significant recall differences across the various IL levels as degree of OL increased. There was no additional information concerning the effects of amount of OL within this period.

PI as a function of List 1 acquisition has been studied by varying the number of trials (Postman & Riley, 1959; Waters, 1942), the number of lists (Underwood, 1945), setting a performance criterion (Atwater, 1953; Underwood, 1949b, 1950a), and analyzing individual item strengths (Runquist, 1957). Two other studies (Greenberg & Underwood, 1950; Werner, 1947) omitted control groups and are not strictly PI designs, and a third (Peixotto, 1947) did not distinguish between learning and retention measures. When significant PI of recall was obtained, all but one of the studies agreed that it was a positive function of the degree or amount of prior learning, and there was even some indication that it leveled off at high degrees of such learning, much as with RI (Atwater, 1953). The one exception (Runquist, 1957) found that PI was not influenced by the degree of the corresponding interfering item strength. The latter is the only study that solely used such analysis and poses an important but separate question concerning the variables determining the retention of

individual items per se. Underwood (1950a) found that PI was eliminated at all degrees of prior learning when recall time was extended to 8-sec. intervals. McGeoch and Underwood (1943), using paired-associates lists, found that, when the pairs were presented in fixed order, thus providing the opportunity for serial learning, significant PI was no longer obtained, as opposed to the usual method in which the order of the pairs is varied. A further indication of the sensitivity of PI to slight procedural changes was given in a report that found significant PI in a serial list at a 2-sec. rate of presentation, but not at a 2.3-sec. rate (Underwood, 1941).

One chronic problem which crops up in studies of the degree of prior learning upon PI (and also in RI designs) is that of controlling for practice and warm-up effects. Traditionally, the control group learns only List 2, whereas the experimental group has had prior practice via List 1. Taking List 2 to a common criterion does not insure equal strengths of learning since the rates of acquisition may differ. Although the problem has been recognized (McGeoch & Irion, 1952), it is not dealt with in most PI studies. Young's (1955) is the only experimental effort at such control, wherein the learning was carried to a seven-eighths criterion on the hypothetical next trial, as determined by previous pilot study data.

The only study of PI as a function of the degree of List 2 learning appeared in the extensive investigation by Postman and Riley (1959) who used serial nonsense lists and naive *S*s. This part of their work revealed a curvilinear PI (both absolute and relative) function. Maximum PI was obtained at the lowest and highest degrees of List 2 acquisition (5 and 40 trials, respectively) across all levels of List 1 training given (5, 10, 20, and 40 trials). Runquist (1957) found that the degree of PI of any individual list item is unaffected by the acquisition strength of that item—again pointing up the discrepancy between single item retention and overall list retention. The study of PI has not kept pace with the growing knowledge about RI, although recently the greater impact of long-range cumulative effects of prior learning have been brought out strikingly by Underwood (1957) who utilized data from previous retention work and showed that more forgetting is attributable to long-range PI effects than to RI. He found that, although well-practiced *S*s forgot about 75% over 24 hours, naive *S*s (no practice lists) forgot only about 30%. This large differential in retention could only be attributed to the strong PI effects of the practice material. Further experimental support was given by Seidel (1959), measuring concurrent PI and RI.

The transitory nature of RI and PI is exemplified in the common observation that these phenomena dissipate after a few relearning trials, sometimes even by the second trial (Osgood, 1948; Underwood, 1945). It follows that recall is the most sensitive measure, whereas if a relearning criterion is used, no interference effects may be demonstrable (McGeoch & Underwood, 1943; Thune & Underwood, 1943; Underwood, 1949b; Waters, 1942).

The rate at which RI dissipates is undoubtedly some function of the degree of learning, or the degree of differentiation of the two response systems involved; but the form of the function is not completely known. Dissipation rate is of importance theoretically and empirically. Melton

and Irwin (1940) obtained fastest dissipation at the highest IL level used (40 trials), followed by the next highest level (20 trials). Thune and Underwood (1943) also found rapid dissipation at the highest levels (10 and 20 trials), but there was no difference in rate between them. This latter finding was incompatible with the two-factor theory of Melton and Irwin, in that it could not be explained by reference to the unlearning factor, because the great differences in overt intrusions obtained under the two conditions should have led to different rates of dissipation, favoring the highest level. This point will be considered again in the section devoted to theory. Data from Underwood (1945, Exp. B) also showed much faster dissipation at the high IL level, and the paper by Briggs (1957) suggests that RI dissipates fastest when the interfering material is well learned or overlearned, only at low and intermediate OL levels. RI persistance was generally found to be greatest at the intermediate IL levels used in the four latter studies. Further data on this point as well as comparable figures for rates of PI would be welcome.

Similarity of Materials

Swenson's (1941) summary of the earlier work on similarity was that "Robinson's theoretical curve is at least roughly accurate" (p. 13). There has since been a definite waning of interest in the Skaggs-Robinson hypothesis as a useful generalization about the effects of similarity upon RI. This is partly because of the failures to duplicate the full theoretical curve within any one experiment (the last attempt at this was made by Kennelly, 1941, and was unsuccessful) and partly because a more heuristic alternative has emerged. The trend within this period may be traced from Boring's (1941) mathematical discussion of communality; Gibson's (1940) more analytical theory reflected in Hamilton's (1943, p. 374) statement that "a two-variable hypothesis should be accepted in preference to the Skaggs-Robinson function"; through Haagen's (1943, p. 44) conclusion that "the hypothesis applies, not to any dimension of similarity, but specifically to the condition in which the continuum of similarity involves a change in the SR relationship of the tasks"; to Osgood's (1949) integration of the literature on RI and similarity in terms of his 3-dimensional transfer and retroaction surface. Ritchie (1954) argued that the Skaggs-Robinson paradox (the statement that the point of maximal OL and IL similarity is simultaneously the condition for greatest interference and also for greatest facilitation) is a pseudoproblem because of an ambiguous scoring procedure. In short, this hypothesis has been superseded by subsequent developments, to which we now turn. Studies of the effects of similarity relationships have been separated into those using paired associates and those using serial lists. The use of paired associates allows specification of the locus of the change in similarity between the lists, an advantage which is not found with serial arrangements. Three classes of change between pair items are possible: response (A–B, A–C), stimulus (A–B, C–B), and both stimulus and response changes (A–B, C–D).

The effect upon retention of learning a new response to an old stimulus has been to produce RI (Bugelski, 1942; Bugelski & Cadwallader, 1956; Gladis & Braun, 1958; Haagen, 1943;

Highland, 1949; Osgood, 1946, 1948; Young, 1955) and, also, retroactive facilitation (Haagen, 1943; Parducci & Knopf, 1958). The variable that determined the direction of the effect was the degree of similarity between the two responses. The problem of developing a rigorously objective quantitative scale of meaningful similarity along dimensions feasible for use in verbal form is a serious one, and it has not been adequately met. Usually, adjectives scaled for varying levels of synonymity to standard words were used. These levels were based upon pooled ratings by judges (Haagen, 1949; Osgood, 1946). Parducci and Knopf (1948) used geometric figures varying along some physical dimension with four-digit numerals varying in identity as the veral responses required. Their OL and recall were visual discrimination tasks, and not really paired associates. The distinction is that the correct response figure and numeral appeared on the stimulus card, whereas in the true paired associates, the response is never a part of the stimulus item. The theoretical rationale of Young's (1955) study deserves some discussion. In the A–B, A–C paradigm, learning A–B also adds to the associative strength of A–C through generalized reinforcement. The magnitude of such generalized reinforcement should be a positive function of the degree of similarity between the B and C response items. In the RI design it was hypothesized that the original list's associative strength (after the IL list was learned) would be the sum of the direct reinforcement gained during its acquisition plus the additional generalized reinforcement gained from the subsequent IL learning. The IL list, on the other hand, would already have gained some generalized reinforcement as a result of the OL training and would thus need less direct reinforcement to achieve criterion during its learning. This would leave the original list with a greater associative strength at recall than the interpolated list, and the magnitude of this difference would be determined by the degree of response similarity between lists. Therefore, it was predicted that, as response similarity between lists increased, RI would decrease and PI would increase. These predictions were tested by Young, using three lists of paired adjectives (to increase the effect) and three levels of response similarity. Results showed that RI as well as overt intrusions decreased as response similarity increased, as predicted. The PI results, as well as a reinterpretation of this entire experiment, will be taken up at the end of this section.

Osgood's (1949) generalization that as response similarity decreases from identity to antagonism, retroactive facilitation should gradually change to increasing RI, was given some empirical support within this period. However, one disturbing finding has emerged. Bugelski and Cadwallader (1956) made a comprehensive attempt to test Osgood's generalizations about similarity effects, part of which involved the use of Osgood's own word lists to define four degrees of response similarity—identical, similar, neutral, and opposed—while keeping the stimuli the same. Results showed decreasing RI with decreasing response similarity. There was more RI with similar than with opposed responses—a finding directly contrary to Osgood's prediction, and not in accord with other data. No explanation was given for these results, but they cast doubt upon the previous formulation of response

similarity. In addition to Osgood's disinclination to use RI control groups, he has also relied upon an uncommon measure of retention, namely, latency scores. In one of his studies (Osgood, 1948), the significant drop in RI between opposed and similar responses was evident only with latency scores, but traditional recall showed no significant differences. In Osgood's other study (1946) there were no significant latency differences at recall, but only on the second and third relearning trials. At no time were the differences between the neutral and opposed conditions significant. All things considered, the evidence in favor of the retroaction surface is less than overwhelming as far as the right half of the response dimension goes, and indicates that a revision is needed.

Saltz (1953) hypothesized that learning A–C after A–B inhibits B. Assuming that inhibition generalizes less than excitation, presenting a slightly altered A stimulus should again tend to evoke B. When tested in a straightforward manner, the hypothesis was not confirmed. A second attempt, designed to minimize changes in set, did result in a tendency toward reappearance of B. No further RI work along these lines has been reported.

There have been two papers on the effects of response similarity on PI. One reported no differential effect (Young, 1955), although overt intrusions increased with response similarity, and the other (Morgan & Underwood, 1950) found that PI tended to decrease as response similarity increased. Osgood (1946, 1948) reported results couched in terms of PI, but his data are for List 2 acquisition and therefore are measuring negative transfer. A methodological oversight with consequent possible confounding of the results of the Young (1955) and Morgan and Underwood (1950) studies should be pointed out. They both varied similarity along the synonymity of meaning dimension. In terms of A–B, A–C, the C response varied from very high (i.e., discreet-ailing, discreet-sickly), to very low similarity, or neutrality with regard to the B response (i.e., noiseless-sincere, noiseless-latent). Each single list had all of the responses at the same similarity level. Thus, it is conceivable that *S* could "catch on" that the List 2 responses were similar in meaning to those of List 1, and thereby reduce his chances of making errors by restricting his responses to members of the synonym category, with a resulting high positive transfer and low apparent PI. This postulated shift in the pool of responses available to *S* could be made entirely without his awareness, as several studies of verbal operant conditioning have demonstrated. With lists of low similarity on the other hand, the possibility of such an occurrence would be nil, and therefore no response class restriction would be made, resulting in a drop in positive transfer and higher apparent PI. Since these studies address themselves to rote learning and retention, the possibility of such a form of concept formation is a serious confounding variable. The test of retention may not be of rote recall at all, but actually of reconstruction of the response on the basis of the general concept of synonymity. As would clearly be predicted by such a "categorization" approach, the learning of List 2 was in fact fastest with high response similarity and became progressively slower with decreasing similarity. Both studies stressed the previously discussed response generalization rationale which would lead to increasing PI with in-

creasing similarity, because learning a similar List 2 response would add to the interfering strength of the List 1 response through generalized or "parasitic" reinforcement. These predictions were not in fact confirmed; rather, PI tended to decrease with increasing similarity (although not statistically significant), an expectation consistent with the categorization hypothesis. The magnitude of the effect is probably dependent upon the relative strength of the two lists, as well as upon the number of alternatives in the response classes, which is a task for further empirical work to verify. Such an unintended source of bias may also have been working in the Bugelski and Cadwallader (1956) study, which used a similar list construction technique. Preferably, items at varying levels of response similarity should be included within the same list, so that *S* would have no opportunity to grasp the concept of the overall list structure. Such a procedure was used for RI by Osgood (1946, 1948) who was aware of this problem. A paper by Twedt and Underwood (1959), which showed that there was no difference in transfer effects between "mixed" and "unmixed" lists, is relevant to lists differing only in formal characteristics, but does not bear upon the question of the general synonymity of the list items as a whole. The lists of the latter study were not varied in degree of meaningful response similarity and thus do not constitute a test of the categorization hypothesis. However, an important paper by Barnes and Underwood (1959) suggested a mediation rationale as another possibility. If A–B is the first list and A–B′ the second, there is a possibility of an A–B–B′ mediation occurring at recall. In view of these complications, we must conclude that the effects of varying response similarity still have not been unequivocally demonstrated or explained.

The retention effect of learning the same response to a new stimulus was reported in four studies, all of which found retroactive facilitation (Bugelski & Cadwallader, 1956; Haagen, 1943; Hamilton, 1943; Highland, 1949). Similarity was varied either by using geometric figures differing in generalizability (originally developed by Gibson, 1941) or meaningful words scaled for synonymity. The results agreed that retroactive facilitation increased with increasing stimulus similarity. The extreme of similarity is identity, and this produces the most facilitation of all since it amounts to continued practice on the original list. At levels of very low similarity there was some inhibition (Haagen, 1943), and according to Hamilton (1943, p. 375): "When the stimulus forms were of 0 degree generalization there was very little difference in retention in conditions with responses identical and with responses different."

No study has ever tested the effects of opposed or antagonistic stimulus relationships while keeping responses the same. Osgood's (1949) retroaction surface does not extend the dimension of stimulus dissimilarity beyond "neutral" or unrelated, although the response dimension does include "antagonistic" relations. The implication is that stimulus opposition is no different in its effects from stimulus neutrality, although no RI evidence is adduced for such a position. It is conceivable, however, that meaningful stimulus opposition or antonymity would actually result in facilitation of recall, based upon a mediation rationale, since such words would be related by *S*'s previous language experience. If response op-

position is expected to differ in effect from response neutrality, then stimulus opposition might also. There are no corresponding paired-associates studies upon the PI effects of stimulus variation.

The effect of changing both the stimulus and response members of the interfering list is concisely stated by Osgood (1949, p. 135): "negative transfer and retroactive inhibition are obtained, the magnitude of both increasing as the stimulus similarity increases." One experiment did not vary stimulus similarity with unrelated responses (Highland, 1949), four studies did vary stimulus similarity with unrelated responses (Gibson, 1941; Haagen, 1943; McClelland & Heath, 1943; Postman, 1958), and another used three degrees of response similarity as well (Bugelski & Cadwallader, 1956). The five latter reports indicate increasing RI with increasing stimulus similarity, and the one study available shows that this holds over all levels of response similarity tested. Two studies from this group will be more fully described since they represent an intriguing departure from the use of the usual physical or meaningful similarity dimension. McClelland and Heath (1943) used as stimulus items for the original and interpolated lists, respectively, a Kent-Rosanoff stimulus word and the most frequent free-association response made to it. Thus an existing prepotent connection was deliberately introduced. Responses were unrelated, and there was no control group. Recall was significantly less under that condition as compared with the case in which there was no association between the stimuli. Since the related words were not similar in appearance or in meaning (e.g., Thirsty-Water) and since a common mediating response could not account for the directionality of the association, the authors concluded that:

> to define the relation between original and interpolated activities which determines the amount of RI, as similarity or as generalization (plain or mediated) is too narrow a conceptionalization, since it does not cover such a learned, uni-directional relation between the two activities as was demonstrated to be of importance here (p. 429).

This study was not carried far enough to prove the point. A third group is needed, for which the related OL and IL stimuli would be interchanged. If this group would display no better recall than the unrelated stimuli group, then the case for the effect of unidirectionality of relationships upon RI would be established. Postman (1958) used geometric figures as OL stimuli. The IL stimuli were either the identical figures, words describing the figures (i.e., "square"), or color names. Responses were unrelated. Both the figure and word groups showed significant RI, with the former having the largest decrement, while the color group did not. These results were explained in terms of the previously learned connections between figures and their names, with formal similarity producing greater interference than mediated equivalence. The influence of unidirectionally prepotent and mediated connections upon forgetting deserves even more attention that it has received. PI is once again slighted, for there are no paired-associates studies concerning both stimulus and response changes.

We turn now to serial list studies, divided into those employing discrete, unconnected items, and those using connected discourse or some approximation thereto. Effects of similarity relations between discrete item lists were reported in three papers which were relatively unrelated as re-

gards their major purposes. Irion (1946) varied the relative serial positions of the original and interpolated adjectives, with some groups learning the identical words for IL, and others learning synonyms. He concluded that similarity of serial position was an effective variable only when identity of meaning was also present. Since several significant differences for IL were reported, we feel that the main variables were confounded with the uncontrolled degree of IL, rendering the results ambiguous. Melton and von Lackum (1941), in a study designed to test an important deduction from the two-factor theory, used two levels of similarity of interpolated items, and found both RI and PI greater under the high similarity condition. Kingsley (1946), with meaningful words, also found poorer retention with interpolated synonyms as opposed to antonyms. Both of the above studies support the generalization that, with serial lists, RI increases with increasing stimulus similarity, along dimensions of both identical elements and meaningfulness.

Ordinary prose or connected discourse has been, until recently, unusually resistant to demonstrable interference effects. Blankenship and Whitely (1941) studied PI of advertising material (a simulated grocer's handbill) as a function of two levels of judged List 2 similarity. Recall after 48 hours showed greater PI for the more similar condition. Their study actually did not vary degrees of similarity of prose, since one of the two lists was nonsense material, and it may be questioned whether a grocer's handbill resembles prose rather than a list of paired associates. Hall (1955) in an RI design, using a completion test, gave 30 sentences for OL, with IL being more sentences varying in two levels of similarity of topic. Results of that, and of a second, unpublished study, both showed no RI. Deese and Hardman (1954) found no RI for connected discourse under conditions of unlimited response time. Ausubel, Robbins, and Blake (1957), using the method of whole presentation, found no RI. The measure of both learning and recall was a recognition test, largely of substance retention. Peairs (1958) did find RI using a recognition procedure; Slamecka (1959), using grouped *S*s, reported that unaided written recall of a short passage was a negative function of the degree of similarity of topic the interfering passage bore to the original passage.

On the whole, these results were rather discouraging about generalizing RI findings from nonsense material to connected discourse and led to the view that prose was not susceptible to RI, or at least to the similarity variable (Miller, 1951, p. 220). We feel, however, that the difficulty was not in the characteristics of connected discourse, but rather in the methods employed. It is noteworthy that all of the above studies employed the less well-controlled techniques of group testing, whole presentation, unlimited recall times, recognition tests, and the like. When, however, connected discourse was presented in the same manner as the traditional serial list, using the serial anticipation method with individually tested *S*s, significant RI was obtained, and it was clearly shown to be a function of degree of OL and IL, as well as of similarity of OL-IL subject matter (Slamecka, 1960a, 1960b). Any presumption of the uniqueness of connected discourse with regard to these variables is no longer tenable, and the door is now open for further exploration of this area.

Errors in recognition and recall of a story were shown to be a function of the interference provided by the interpolated presentation of a picture which bore some thematic resemblance to the story (Davis & Sinha, 1950a, 1950b). Similarly, Belbin (1950) showed that an interpolated recall test concerning an incidentally present poster interfered with the subsequent recognition of the poster. If the attempted recall is viewed as interfering with the original perceptual trace, then the degree of OL and IL (recall test) similarity was determined by each *S*'s own recall performance.

Lying somewhere between the use of discrete, unconnected items and ordinary prose are two studies employing lists of various orders of approximation to English, constructed according to a method developed by Miller and Selfridge (1950). If RI is a function of contextual constraint, then the use of such materials should be appropriate.

Heise (1956) used an unrelated word list as OL, and five different IL levels of approximation to English. He found recall was best with the greatest dissimilarity between the lists. Thus, the seventh order IL list (close to English text) produced almost no interference, whereas the first order list (same order as OL) produced a great deal, again supporting the generalization concerning greater RI with greater similarity between serial lists. King and Cofer (1958) extended this technique by using OL lists at the zero, first, third, and fifth orders, with four different orders of IL at each of the OL levels. Their intent was to examine similarity effects at various levels of contextual constraint, but the results did not show an overall comprehensive pattern for RI. They suggested that the effects of contextual constraint may prove to be more complex than originally expected, and called for further investigation.

Extrinsic Factors

In this section are papers focusing upon variables actually extrinsic to the specific items being learned. In most of these studies the groups learned identical materials, and they differed only with regard to such things as the general surround, testing methods, and sets.

The striking effects of altered environment were shown by Bilodeau and Schlosberg (1951). The two groups differed only in the conditions under which IL took place. One group stayed in the same room for all phases, and the other had the IL in a dissimilar room with a different exposure device and a changed posture for *S*. Recall, done in the OL room, indicated that IL interfered only half as much when associated with a different surround. Elaborating upon this, Greenspoon and Ranyard (1957) also used two different surrounds (different rooms, posture, and exposure devices designated as A and B), in four combinations, and the results, in terms of decreasing order of recall were ABA (AAA, ABB) AAB (those within parentheses not significantly different). Although no controls were used, the findings agree with those of Bilodeau and Schlosberg. These studies support the view that, since recall takes place in some context, the cues governing a response lie not only within the learning material, but also in the general surround, and that the magnitude of RI is a partial function of such context-carried cues. The relative importance of the proprioceptive vs. the exteroceptive cues was not assessed.

Jenkins and Postman (1949) varied testing procedures for OL and IL, using anticipation (A) or recognition (R), in four combinations. Results showed a significant increase in recall when procedures were different, under only one of the comparisons (A–A, A–R). The authors concluded that using a different testing method is a change in set and "helps in the functional isolation of materials learned successively" (p. 72). Postman and Postman (1948) gave four groups the same materials, differing only in the order of the S–R items. Paired syllables-numbers for OL were followed by either paired numbers-syllables or more syllables-numbers. The changed set groups showed better recall. No control groups were used. In the second part of the same report, OL was paired words with either a compatible (doctor-heal) or incompatible (war-peaceful) relation between them. For IL, half the *S*s learned a list with the same logical relations, and half learned one with the opposite relations to OL. This latter group showed superior retention, again attributed to the dissimilar sets involved.

Comparing the effects of incidental vs. intentional learning of OL and IL, Postman and Adams (1956) found that, regardless of the OL conditions, intentional IL produced more RI than incidental IL. Both intentional and incidental learning were equally susceptible to RI when followed by IL of the same kind and strength as OL. The authors noted that: "Intentional practice resulted in the learning of a longer number of items during interpolation and hence was a more effective source of interference" (p. 328). Thus, it appears that these conditions were simply the vehicles by which degree of IL, the effective variable, was manipulated. In an earlier paper, Prentice (1943) concluded that incidental learning was more subject to RI than intentional, but when Postman and Adams (1956) corrected Prentice's data by subtracting the respective control group scores, the results agreed with the Postman and Adams findings. If incidental and intentional conditions are construed as providing different sets, or "functional isolation," then an experiment in which the degree of acquisition was equalized should be expected to give different results: the similarly treated groups should display more RI than the changed-set groups. Since this has not been done, we must conclude that the RI effects of incidental vs. intentional conditions per se are not yet known.

The effect of the emotion-arousing characteristic of the IL upon retention is an interesting question, but only one study attempted it within this period and produced inconclusive results (McMullin, 1942), probably because of a confounded experimental design. Among the truly inherent subject variables that have been investigated is the effect of the age of *S* (Gladis & Braun, 1958; Wywrocki, 1957). The former study divided *S*s into three age classes: 20–29, 40–49, and 60–72 years. There was no control group. Although a negative relationship between age and rate of learning was found, the adjusted absolute recall scores revealed no differential RI effects related to age. One might speculate that the decreased learning ability of the older *S*s was a PI effect resulting from their many years of previous learning. When the recall scores were "corrected" for this, the actually obtained negative relation between raw recall and age was eliminated. Among the more clinical subject

variables, Cassel (1957) reported no differential RI susceptibility between *S*s of normal mentality and those with mental deficiency. Sherman (1957) found that psychopaths showed better retention than either neurotics or normals, measured by total forgetting scores. Livson and Krech (1955) reported a moderate positive correlation between recall and scores on the KAE (Kinesthetic Aftereffect Test, which was related to Krech's cortical conductivity hypothesis).

The importance of set factors, generally called warm-up effects, has been recognized (Irion, 1948). Thune (1958) showed that recall was significantly facilitated by a preceding appropriate warm-up. If OL was from a memory drum and IL from a filmstrip, then a memory drum warm-up facilitated recall, but a filmstrip warm-up did not. Inappropriate warm-up did facilitate later relearning trials, and Thune concluded that warm-up has both peripheral and central components, with the former more transitory. No RI control groups were used.

The effects of such extrinsic variables upon PI have not yet been investigated. This line of research should be extended, since the magnitudes of interference obtained are often considerable, and probably much of our everyday forgetting is attributable to such context-associated factors.

Temporal Effects

Swenson (1941) summarized the effects of temporal variables as follows:

> [*a*] . . . interpolation immediately adjacent either to original learning or to recall of original learning is more effective in producing retroactive inhibition than is interpolated activity between those two extremes (p. 15). [*b*] . . . the more recent studies suggest an inverse relationship between length of the time interval and relative retroactive inhibition (p. 16).

Subsequent work has called for a modification of those statements.

Examination of the RI paradigm reveals three manipulable temporal intervals: end of OL—start of IL, end of IL—start of RL, and end of OL—start of RL. No single experiment, while keeping the IL learning period constant, can vary only one of these intervals without automatically changing one of the others. When the IL learning period varies (as in studies giving different numbers of IL trials) while the OL–IL and the OL–RL intervals are kept constant, then the IL–RL interval will inevitably vary. Therefore, in the study of any one of these variables, confounding is inescapable. There is no easy way out of this dilemma. The only technique approaching a solution seems to be to do several separate experiments, confounding a different pair of intervals each time, and then evaluating the results of all the experiments by determining which confoundings have no effect. This more elaborate approach has not been used in actual practice; rather, acceptance of such confounding seems to be the rule.

Varying the IL–RL interval allows for measurement of progressive changes in the strength of RI and PI, and deductions concerning the events that occur in that time. Underwood (1948a), using IL–RL intervals of 5 and 48 hr., and Briggs (1954) at 4 min. to 72 hr., report no significant changes in magnitude of RI. Deese and Marder (1957), using unlimited response times, from intervals of 4 min. to 48 hrs., and Peterson and Peterson (1957) from 0 to 15 min., both found no changes in recall. Slight RI decreases were reported by

Jones (1953) from .17 to 24 hrs. (with an increase from 24 to 144 hrs.) and by Ishihara (1951). Using the uncommon A–B, C–D design with very high levels of practice, Rothkopf (1957) found an increase in recall from 0 to 21 hrs., but no control groups were used. From the trend of these results, the best conclusion seems to be that RI remains relatively stable over time, at least up to 72 hrs.

In examining the temporal course of PI, Underwood (1949b) found no change from 20 to 75 min., but (Underwood, 1948a) did find a drop in recall from 5 to 48 hrs. (no control groups), and Jones (1953) also reported increasing PI. In a study not explicitly designed to assess PI, therefore lacking control groups, Greenberg and Underwood (1950) also found a significant drop in List 2 recall from 10 min. to 5 hrs. to 48 hrs. In spite of the lack of appropriate controls in some of these studies, the results are in sufficient agreement to allow the conclusion that PI shows a gradual increase through time, which is in accord with logical expectations, as Underwood (1948a) has pointed out.

In comparing the relative strengths of RI vs. PI through time under comparable conditions, Underwood (1948a) found that RI was greater at 5 hrs., but that there was no difference at 48 hrs. Jones (1953) and Rothkopf (1957) reported similar observations. Underwood hypothesized that the failure of List 1 recall to diminish might be due to a process of gradual recovery of OL responses after their unlearning during IL. This led to the use of the modified free recall (MFR) procedure as a method of assessing response dominance. In MFR, *S* is given a stimulus item common to both lists and asked for the first response that comes to mind. It was felt that such unrestricted, uncorrected recall would provide a fairer estimate of the relative strengths of the competing responses, although it was clearly not intended to be equivalent to the restricted recall required for RI measures. Underwood (1948b) gave MFR at 1 min., 5, 24, and 48 hrs. after IL and found no change in OL responses, a consistent drop in IL responses, and a rise in "other" responses. He concluded that:

> These data are given as further support of the interpretation of unlearning of the first list as being similar to experimental extinction. The fact that no decrease in the effective strength of the first list responses takes place over 48 hrs. suggests that a process running counter to the usual forgetting process is present. It is suggested that this mechanism may be likened to spontaneous recovery (p. 438).

Concerning OL responses, it seems unnecessary to hypothesize two opposing tendencies (recovery vs. "usual forgetting") canceling each other out, as it were, to account for a finding of no change. The usual forgetting curve might not necessarily be expected of OL responses, since the effects of IL could be such as to obliterate, through differential unlearning, more of the weak than the strong responses, leaving the strong, stable ones that are more resistant to the "usual forgetting" process, in the preponderance. List 2 responses, not so selectively eliminated, would be expected to decrease in time. In support of this alternate view we call attention to two relevant bits of evidence. Deese and Marder (1957) found that the number of items recalled after interpolation remained constant over intervals of 4 min., 2, 24, and 48 hours after IL. Also, Runquist (1957) found that resistance to RI was positively related to

the degree of an original item's strength. In another MFR experiment, Briggs (1954) did obtain a rise in OL responses between 4 min. and 6 hrs., with subsequent stability through 72 hrs. Because of the discrepancy between these data and those of Underwood (1948b), another study was done in which Briggs, Thompson, and Brogden (1954) found no OL changes between 4 min. and 6 hrs. These authors concluded that "responses from original learning show no change, that responses from interpolated learning tend to decrease with time interval in a fairly regular manner, and that 'other' responses tend to increase . . . " (p. 423). From these MFR data, we tentatively conclude that the processes underlying the temporal stability of RI do not as yet clearly indicate the recovery of unlearned original responses. The spontaneous recovery hypothesis is an attractive one, but more evidence of its validity should be brought forth.

Another problem of interest is the effect of the temporal point of interpolation, which requires keeping the OL–RL interval constant and varying the OL–IL period. Unavoidably, this introduces confounding with the simultaneously varying OL–IL interval, as discussed above.

Houlahan (1941) gave IL either 0, 4, or 8 mins. after OL and found more RI for the immediate interpolation condition. However, there was no direct measure of OL; rather, the performance on some previously learned lists thought to be of equal difficulty to OL was used as a comparison. Within a 16-day OL–RL period, Postman and Alper (1946) gave IL at eight evenly dispersed intervals and found maxima of recall at 1, 8, and 15 days after OL. Degree of acquisition was uncontrolled, since fixed numbers of trials were given; and, since no acquisition data were presented, unequivocal conclusions about the temporal variable cannot be drawn.

Maeda (1951), using short intervals, reported greatest reproduction when IL directly followed OL. Newton (1955) with an A–B, C–B design, and Archer and Underwood (1951) with A–B, A–C, using a 48-hr. OL–RL period with IL at 0, 24, and 48 hrs., concluded that temporal point of interpolation was not an effective variable. Newton and Wickens (1956) noted that the Archer and Underwood study failed to control for differential warm-up, in that the group with IL immediately before RL benefitted by warm-up, whereas the other two groups had no comparable advantage. They repeated the Archer and Underwood study with the same materials, but gave a warm-up task to the 0- and 24-hr. groups. No effects of the temporal intervals were obtained, confirming the previous results. However, they also reported two additional experiments, with an A–B, C–D design, with warm-up provided. One study had a performance acquisition criterion, and the other a fixed number of trials. Results of both showed that the 48-hr. group did show significantly more RI than the other two. Those authors state that the A–B, A–C design "is a relationship which is designed to produce a maximum amount of RI, and the intensity of this condition may obscure the RI which can arise from a variable of lesser importance—as the temporal variable may well be" (Newton & Wickens, 1956, p. 153). They especially stressed the importance of generalized competition between lists, a point which shall be developed

further in the theoretical section below. We tentatively conclude from the Newton and Wickens (1956) data, supported by Maeda (1951), that RI increases as the OL–IL interval increases and that the effect is thus far specific to the A–B, C–D design.

A comparable PI design would require a constant List 2–RL period, while varying the List 1–List 2 interval. We have been unable to find such an experiment in the literature within this period. Ray (1945) studied List 2 acquisition as a function of the interval since the learning of List 1. Although he speaks of PI, the design is appropriate only to conclusions about negative transfer.

Another temporal variable which has not been studied sufficiently in an RI design is the rate of presentation of the items. The only relevant retention study on this is an unpublished honors thesis by Seeler (1958). OL was a 35-word passage of prose presented via tape recording to a criterion of one perfect unaided written recall, followed by similar memorization of an IL passage, and then by OL recall. OL rates of presentation were ½, 1, and 2 secs., followed by ½; 1; or 2-second counterbalanced rates on the IL. No control groups were used. Results showed that number of trials to mastery of all original and interpolated passages was a direct function of their presentation rates, a finding consistent with acquisition reports based on nonsense materials. There was no influence of either the OL or IL presentation rates, or any of their combinations, upon recall. It might have been supposed that an IL rate different from the OL rate would have served to functionally isolate the original list and produce less forgetting, but that was not the case. The possibility of confounding rate of presentation and strength of associations at the end of OL due to differential acquisition rates (Underwood, 1949a) is not a problem in this study, since unaided recall was used. With the method of serial anticipation, however, the criterial OL trial is also another learning trial, and two groups taken to the same performance criterion may still differ on total associative strength at the termination of the last OL trial. The problem is always present whenever any variable that effects rate of acquisition (such as meaningfulness, similarity, etc.) is used along with the serial anticipation technique. The generalizability of the latter results to unconnected materials as well as the additional independent problem of the RI effects of massed vs. distributed training must await further study.

Major Theoretical Positions

In this section we shall discuss the four main theoretical positions which have influenced the period covered by this review. Two major formulations, appearing within a few months of each other (Gibson, 1940; Melton & Irwin, 1940), guided the theoretical aspects of the study of RI within the first few years covered by this review.

Utilizing the classical conditioning principles of stimulus generalization and differentiation, Gibson (1940) presented a set of postulates for verbal behavior that served to lend greater predictive specificity to the transfer or straight competition-of-response view, previously developed by McGeoch and his collaborators. Basic to Gibson's approach is the view that verbal learning and retention are matters of developing discriminations among the items to be learned. She defines her two basic

constructs as follows: The construct of generalization is "the tendency for a response R_a learned to S_a to occur when S_b (with which it has not been previously associated) is presented" (p. 204). The construct of differentiation is "a progressive decrease in generalization as a result of reinforced practice with S_a–R_a and reinforced presentation of S_b" (p. 205). A curvilinear growth function of the generalization tendency as practice trials increase is stressed. Essentially, RI is related to the degree of discriminability of the two lists, such discriminability being a positive function of their respective degrees of learning, and a negative function of the time elapsed since learning. Spontaneous recovery of generalization tendencies (wrong responses) through time is assumed. From these postulates, several deductions concerning RI were presented, and some of these have been tested and confirmed: for instance, RI as a function of various similarity relations among the items (Gibson, 1941; Hamilton, 1943), and the curvilinear RI function obtained as the degree of IL increases (Melton & Irwin, 1940). Among the deductions tested but not confirmed is one bearing upon the temporal point of interpolation problem. Gibson feels that one of the reasons for the disparity of results on this question lies in the neglect of the importance of the degree of acquisition of the lists. She predicted that acquisition level would be found to interact with the temporal point of interpolation because the spontaneous recovery of generalization tendencies between lists is a function of time. This prediction was tested by Archer and Underwood (1951) using three levels of IL acquisition (6/10, 10/10, and 10/10+5 trials) and three OL–IL intervals (0, 24, and 48 hrs.), but no interaction between them was found. RI control groups were not used, and in light of the theoretical importance of this study it would seem advisable to re-examine these variables with a design adaptable to relative RI measures. The authors themselves expressed dissatisfaction with the outcome and "felt that a modification of the conditions in our design would indicate the temporal position to be a factor" (p. 289).

Considering the general reaction toward Gibson's theory in succeeding RI work, we feel that, on the whole, it has been favorably received, since it has been given a certain amount of implicit corroboration by way of being compatible with many findings (for instance, Briggs, 1957) and has potential for even further development. It has not, however, stimulated a comprehensive series of experiments aimed at testing the many RI deductions implicit within it. The reason for this is certainly not any lack of clarity in the postulates. One present weakness seems to be the lack of direct evidence for a spontaneous recovery process influencing RI.

Melton and Irwin (1940) introduced their two-factor theory within the framework of a study of RI as a function of the degree of IL. OL was 5 trials on an 18-item serial nonsense list, followed by 5, 10, 20, or 40 trials on an IL list. Relying upon a count of the overt interlist intrusions as an objective index of the degree of competition between original and interpolated responses at recall, they found that the curves of amount of absolute RI, and the number of such intrusions (multiplied by a factor of 2 to do justice to partial intrusions) were not highly correlated. (The theoretical importance of intrusion counts gained its ascendancy with

this study.) Rather, interlist intrusions increased to a maximum at intermediate IL levels and then decreased markedly, whereas the curve of RI rose sharply and maintained a relatively high level, declining slightly at the highest degree of IL. That portion of the RI attributable to direct competition of responses at recall was at a maximum when OL and IL were about equal in strength. Therefore, to account for the remainder of the obtained RI not accounted for by overt competition, Melton and Irwin postulated another factor at work, tentatively identified as the direct "unlearning" of the original responses by their unreinforced elicitation or punishment, during IL. The growth of this "Factor X" was assumed to be a progressively increasing function of IL strength. Since Factor X was almost totally responsible for the absolute RI at the highest IL level, and since RI under that condition dissipated most rapidly after a few relearning trials, it was concluded that the effects of such unlearning were quite transitory. This was still a competition of response theory in the sense that the original responses were still assumed to be competing at recall with the interpolated ones, but to that was added the factor of weakening in OL response strength, if not complete extinction, through the process of unlearning.

The presence of confounding between the degree of IL, and the end of IL–start of RL interval was pointed out by Peterson and Peterson (1957) as a possible alternative account of the differences in intrusions obtained by the Melton and Irwin design. With a fixed OL–RL interval the IL–RL interval shortens, with increasing IL trials taking more time. However, another study of the effects of degree of IL did use a fixed IL–RL interval (with a correspondingly varying OL–RL interval—Osgood, 1948) and still found comparable intrusion changes.

A direct deduction from the two-factor theory is that RI, being a result of both unlearning and competition effects, should be greater than PI, which was presumed to be the result of response competition alone. This hypothesis was tested and confirmed by Melton and von Lackum (1941) in a study using five trials on each of two 10-item consonant lists, and has also been given further general support by others (Jones, 1953; McGeoch & Underwood, 1943; Underwood, 1942, 1945). Underwood (1948a) in yet another study also found greater RI than PI at 5 hrs.; but at 24 hrs. they were equal. His resulting postulation of spontaneous recovery of the OL, and the subsequent developments of that concept have been discussed above.

Later, certain other observations led to some discontent with the two-factor theory. In an experiment designed to test the generalizability of the Melton and Irwin findings to paired-adjectives lists, Thune and Underwood (1943) used an A–B, A–C design with five OL trials and 0, 5, 10, or 20 IL trials. Their results confirmed the existence of a negatively accelerated function between RI and degree of IL, as well as the fact that overt intrusions were maximal at the intermediate IL levels (10 trials) and declined sharply by the 20-trial level, while RI still remained massive. However, there was no difference in the rate of RI dissipation between the 10 and 20 trial IL levels, and therefore the transitoriness of RI at these levels could not reasonably be attributable to the unlearning construct. The two-factor

theory would have been forced to predict faster dissipation at the 20 IL level, since overt intrusions were far less for it than for the 10 IL level. In addition, the curve of Factor X drawn for the Thune and Underwood data was quite different in shape from that obtained by Melton and Irwin, and it was felt to imply rather incongruous psychological properties for a curve of unlearning. In addition, an item analysis revealed that almost half of the overt intrusions took place on items where the original response had never been reinforced (or correctly anticipated) at all! Therefore, such interlist intrusions could not be legitimate indicators of response competition, since those responses had never been learned during OL, and were simply not available to be competing with anything. It is also to be expected that for original responses to be unlearned they would have to occur during IL in sufficient frequency to be subject to punishment or lack of reinforcement. Yet, as Osgood (1948) pointed out from his data, the number of related original list intrusions during IL was "infinitesimally small" and could not possibly account for much unlearning at all. This previously observed discrepancy between the assumed growth of Factor X and the lack of increase in intrusions during IL as a function of increasing IL trials should be tempered with the possibility that partial intrusions could still play a large role in determining the degree of unlearning obtained, and such intrusions are not easily detected and counted.

Thune and Underwood (1943) suggested that the ratio of overt to covert (and partial) errors need not necessarily remain constant, but may undergo progressive change as a function of the degree of IL, therefore accounting for the drop in overt intrusions by postulating an increase in implicit interference. In a subsequent paper Underwood (1945) elaborated upon this suggestion and formalized his differentiation theory.

The shift in error ratios was interpreted as a resultant of two simultaneous processes: increasing IL associative strength tending to produce more overt intrusions, but being gradually overcome by the growth of differentiation, tending to reduce the intrusions. The magnitude of the differentiation construct was held to be a positive function of the degree of learning of both lists and a negative function of the time between the end of IL and the start of RL. A decrease in overt intrusions was, in effect, the index of increasing differentiation. When the two lists are about equally well learned, intrusions are maximal and differentiation is low; but with increasing disparity between their absolute or relative acquisition levels, intrusions are reduced, indicating increased differentiation. By the same token, a short IL–RL interval should also produce higher differentiation. That this is in fact the case was shown in the Archer and Underwood study (1958) where overt intrusions declined as the IL–RL interval became shorter. The increasing differentiation allows *S* to recognize and withhold erroneous responses, resulting in fewer interlist intrusions and more covert or omission errors. Differentiation was described phenomenologically by Underwood (1945, p. 25) as being

> related to the verbally reported experience of "knowing" on the part of the subject that the responses from the interpolated learning are inappropriate at the attempted recall of the OL. Degree of differentiation in this sense is thus an indication of the degree to which the subject identifies the list to which each response belongs.

Empirical support for various aspects of this theory has come from several studies (e.g., Archer & Underwood, 1951; Osgood, 1948; Thune & Underwood, 1943; Underwood, 1945). Further, the fact that intrusion frequencies change but RI still remains constant might be simply a function of the limited recall time (usually 2 sec.) available to *S*. If this recall time was extended, then perhaps *S* would have sufficient time both to recognize the erroneous and verbalize the correct response, thus displaying a decrease in RI at high IL (differentiation) levels. Underwood (1950a, 1950b) tested this promising hypothesis, but found no dropping off of the Melton and Irwin effect, and concluded that differentiation does not change as a function of increased (8 sec.) recall time. Unlearning was therefore still retained as a useful concept; but, since it was shown that such response weakening took place only in the first "few" IL trials, as measured by associative inhibition, and because of the relatively great stress put upon the role of differentiation, Underwood's revision of the two-factor theory became an important independent influence upon subsequent RI thinking.

Certain apparent similarities between Underwood's differentiation construct and Gibson's concept of differentiation deserve to be pointed out at this time. For both theorists, differentiation is in part a positive function of degree of reinforced practice on the material, such practice serving to reduce overt intrusion errors. Secondly, temporal relationships also play a large part in determining the strength of both constructs. However, the two positions do differ with regard to certain important aspects of operation of these determiners of differentiation. Underwood's concept refers to the more global process of *S* correctly assigning the list membership of the responses, whereas Gibson speaks of discrete S–R connections in competition. Furthermore, Underwood's theory is derived from experiments based largely upon the A–B, A–C design, whereas for Gibson, generalization as defined requires that the stimulus members be similar, but not identical. For Underwood, increasing differentiation is marked by a reduction of intrusions and an increase in omissions, but no drop in RI, whereas Gibson implies that increasing differentiation will result directly in improved performance. And finally, Gibson makes spontaneous recovery an integral part of her differentiation concept, while it was not until later that Underwood suggested a spontaneous recovery process, and that was reserved for the unlearning aspect of his theory.

The last theoretical formulation to be considered was put forth by Osgood (1946). It stemmed from his investigations of the RI effects of meaningful opposed responses and involved a hypothesis about reciprocal inhibition of antagonistic reactions, wherein "simultaneous with learning any response the *S* is also learning not to make the directly antagonistic response" (Osgood, 1948, p. 150). This was clearly an application of the reciprocal inhibition concept of neurophysiology to the area of verbal behavior. In pursuing the tenability of this position, two relevant transfer studies have shown that the learning of both similar and opposed List 2 responses was equally rapid, and much easier than learning neutral responses (Ishihara & Kasha, 1953; Ishihara, Morimoto, Kasha, & Kubo, 1957), thus failing to confirm the hypothesis. Unless fur-

ther support for the hypothesis is forthcoming, we must conclude that it will not become an important influence in RI work.

With regard to the question of the adequacy of the two-factor theories we are of the opinion that the concept of unlearning is a valuable one, but that an acceptable measure of its magnitude has not yet been devised. Interlist intrusions were proposed only as a partial index, but not as a complete measure of its effects, and the difficulties encountered by such an index have been enumerated above. Instructions calculated to encourage the verbalizing of errors do just that: Morrow (1954) and Bugelski (1948) found that "all that is required to obtain a large number of such errors is to ask for them" (p. 680).

Two interesting proposals have been advanced as methods for distinguishing operationally between effects of competition and effects of unlearning. Postman and Kaplan (1947) spoke of two measures of RI: error scores, and the reaction times for correct responses (residual retroaction). These two measures were found not to be correlated and are thus of necessity measures of two different processes. They suggest that: "It is possible that retention loss (error scores) reflects the effects of unlearning, whereas reaction times may depend primarily on the competition between responses" (p. 143). Their experiment did not include variation of any factor which might be expected to affect unlearning differentially and therefore the usefulness of their proposal has not yet been tested.

Later, Postman and Egan (1948) proposed that the rate of recall of correct responses be a measure of unlearning. Retention was measured by the free recall procedure, and performance was recorded both in terms of number of items recalled, as well as by the rate of emission of correct items, per 3-sec. periods. They state that

> The two types of measures—amount lost and rate of recall—may be regarded as measures of these two processes (unlearning and competition, respectively). Those aspects of OL which have been unlearned cannot be evoked on retest: unlearning leads to decrement in amount retained. Other aspects suffer competition from the IL but are not unlearned. They are potentially available but "disturbed," and manifest that in a slower rate of recall (p. 543).

These are both valid and constructive formulations deserving of further attention, but no significant efforts have as yet been made to test their usefulness in predicting data crucially relevant to the unlearning factor.

These experiments by Postman point in a new direction, suggesting that such evidence for competition of response is a result of the brief recall times used in RI studies. If competition of response results in increased latencies, then decrements in recall may come when the latency of a response exceeds the 2-sec. interval usually used. Underwood's (1950a) study, which found no PI with an 8-sec. recall interval, supports this possibility.

An experiment by Ceraso (1959) may provide further support for such a hypothesis. With an A–B, A–C design, *S*s were asked to recall both the first and second list responses and also to assign these to the proper list. Since a 20-sec. (maximum) recall interval was used, blocking due to competition of response should not be expected. An analysis of the first list responses which were correct on the last trial of OL, and were then scored as incorrect at recall, showed that the reason for the forgetting was

simply the unavailability of the response. If the response was available at recall, it was also assigned to the correct list. Since competition of response should reveal itself as a misassignment of the response, it was clear that the forgetting obtained could not be accounted for by competition. Using a technique somewhat related, with an A–B, A–C design, Barnes and Underwood (1959) obtained similar results, and accordingly rejected a competition explanation.

Ceraso also found that in a large number of cases *S* could give both responses to the stimulus. But does not the unlearning hypothesis imply that learning the second list response entails the unavailability of the first list response? The answer that immediately suggests itself is that unlearning is a function of the degree of first and second list item learning. Therefore, an item analysis of the kind performed by Runquist was undertaken. The result showed that degree of learning of the second list item did not affect the retention of the first list item, thus verifying Runquist's (1957) original finding.

It seems that the latter data pose a real problem for current theories of RI, since the basic mechanism usually postulated requires interaction between associations with similar or identical stimulus items. Both the Runquist and Ceraso findings seem to indicate a nonspecific mechanism. Learning a second list affects the entire first list, regardless of the specific item interactions.

In conclusion, it appears that the major theoretical accounts of RI have remained relatively unchallenged and unchanged for the last ten years, in spite of the accumulation of considerable empirical data. It is hoped that this overview of the current state of the field will help to initiate a more vigorous and sustained effort toward an improved theory of forgetting.

Summary

For a concluding statement we feel it would be appropriate to enumerate some of the pressing problems and empirical gaps currently evident in the status of our knowledge of RI and PI. These points are presented in the order of their appearance in the foregoing review and do not reflect any opinion regarding their relative importance.

1. Reconsideration of the relative merits of RI quantification: absolute RI, relative RI, and total forgetting
2. Determinants of the RI and PI of individual items
3. Determinants of the rate of PI dissipation
4. Development of an objectively quantitative scale of similarity for use in constructing lists of items
5. Reappraisal of the right half of the response dimension of Osgood's retroaction surface
6. Effects of opposed or antagonistic stimulus relations upon RI and PI, with responses the same
7. Effects of varying response similarity upon PI, with the "categorization approach" error eliminated
8. Further study of the RI effects of mediated and unidirectional prepotent association between list items
9. PI as a function of similarity relations within the A–B, C–D design
10. Determinants of the RI and PI of connected discourse
11. Relative importance of proprioceptive vs. exteroceptive extrinsic cues for recall
12. RI effects of incidental vs. intentional acquisition conditions, with degree of acquisition controlled

ther support for the hypothesis is forthcoming, we must conclude that it will not become an important influence in RI work.

With regard to the question of the adequacy of the two-factor theories we are of the opinion that the concept of unlearning is a valuable one, but that an acceptable measure of its magnitude has not yet been devised. Interlist intrusions were proposed only as a partial index, but not as a complete measure of its effects, and the difficulties encountered by such an index have been enumerated above. Instructions calculated to encourage the verbalizing of errors do just that: Morrow (1954) and Bugelski (1948) found that "all that is required to obtain a large number of such errors is to ask for them" (p. 680).

Two interesting proposals have been advanced as methods for distinguishing operationally between effects of competition and effects of unlearning. Postman and Kaplan (1947) spoke of two measures of RI: error scores, and the reaction times for correct responses (residual retroaction). These two measures were found not to be correlated and are thus of necessity measures of two different processes. They suggest that: "It is possible that retention loss (error scores) reflects the effects of unlearning, whereas reaction times may depend primarily on the competition between responses" (p. 143). Their experiment did not include variation of any factor which might be expected to affect unlearning differentially and therefore the usefulness of their proposal has not yet been tested.

Later, Postman and Egan (1948) proposed that the rate of recall of correct responses be a measure of unlearning. Retention was measured by the free recall procedure, and performance was recorded both in terms of number of items recalled, as well as by the rate of emission of correct items, per 3-sec. periods. They state that

> The two types of measures—amount lost and rate of recall—may be regarded as measures of these two processes (unlearning and competition, respectively). Those aspects of OL which have been unlearned cannot be evoked on retest: unlearning leads to decrement in amount retained. Other aspects suffer competition from the IL but are not unlearned. They are potentially available but "disturbed," and manifest that in a slower rate of recall (p. 543).

These are both valid and constructive formulations deserving of further attention, but no significant efforts have as yet been made to test their usefulness in predicting data crucially relevant to the unlearning factor.

These experiments by Postman point in a new direction, suggesting that such evidence for competition of response is a result of the brief recall times used in RI studies. If competition of response results in increased latencies, then decrements in recall may come when the latency of a response exceeds the 2-sec. interval usually used. Underwood's (1950a) study, which found no PI with an 8-sec. recall interval, supports this possibility.

An experiment by Ceraso (1959) may provide further support for such a hypothesis. With an A–B, A–C design, *S*s were asked to recall both the first and second list responses and also to assign these to the proper list. Since a 20-sec. (maximum) recall interval was used, blocking due to competition of response should not be expected. An analysis of the first list responses which were correct on the last trial of OL, and were then scored as incorrect at recall, showed that the reason for the forgetting was

simply the unavailability of the response. If the response was available at recall, it was also assigned to the correct list. Since competition of response should reveal itself as a misassignment of the response, it was clear that the forgetting obtained could not be accounted for by competition. Using a technique somewhat related, with an A–B, A–C design, Barnes and Underwood (1959) obtained similar results, and accordingly rejected a competition explanation.

Ceraso also found that in a large number of cases *S* could give both responses to the stimulus. But does not the unlearning hypothesis imply that learning the second list response entails the unavailability of the first list response? The answer that immediately suggests itself is that unlearning is a function of the degree of first and second list item learning. Therefore, an item analysis of the kind performed by Runquist was undertaken. The result showed that degree of learning of the second list item did not affect the retention of the first list item, thus verifying Runquist's (1957) original finding.

It seems that the latter data pose a real problem for current theories of RI, since the basic mechanism usually postulated requires interaction between associations with similar or identical stimulus items. Both the Runquist and Ceraso findings seem to indicate a nonspecific mechanism. Learning a second list affects the entire first list, regardless of the specific item interactions.

In conclusion, it appears that the major theoretical accounts of RI have remained relatively unchallenged and unchanged for the last ten years, in spite of the accumulation of considerable empirical data. It is hoped that this overview of the current state of the field will help to initiate a more vigorous and sustained effort toward an improved theory of forgetting.

Summary

For a concluding statement we feel it would be appropriate to enumerate some of the pressing problems and empirical gaps currently evident in the status of our knowledge of RI and PI. These points are presented in the order of their appearance in the foregoing review and do not reflect any opinion regarding their relative importance.

1. Reconsideration of the relative merits of RI quantification: absolute RI, relative RI, and total forgetting
2. Determinants of the RI and PI of individual items
3. Determinants of the rate of PI dissipation
4. Development of an objectively quantitative scale of similarity for use in constructing lists of items
5. Reappraisal of the right half of the response dimension of Osgood's retroaction surface
6. Effects of opposed or antagonistic stimulus relations upon RI and PI, with responses the same
7. Effects of varying response similarity upon PI, with the "categorization approach" error eliminated
8. Further study of the RI effects of mediated and unidirectional prepotent association between list items
9. PI as a function of similarity relations within the A–B, C–D design
10. Determinants of the RI and PI of connected discourse
11. Relative importance of proprioceptive vs. exteroceptive extrinsic cues for recall
12. RI effects of incidental vs. intentional acquisition conditions, with degree of acquisition controlled

13. Effects of the affective characteristics of the material upon its RI and PI

14. Better handling of the problem of confounding which arises when temporal intervals are manipulated

15. Further tests of the validity of the spontaneous recovery hypothesis

16. Examination of the point of interpolation problem as a function of other attendant variables

17. RI as a function of presentation rate, and of massing vs. distributing trials

18. Testing of the two-factor theory through an improved measure of the unlearning construct

REFERENCES

Archer, E. J., & Underwood, B. J. Retroactive inhibition of verbal association as a multiple function of temporal point of interpolation and degree of interpolated learning. *J. exp. Psychol.*, 1951, **42**, 283–290.

Atwater, S. K. Proactive inhibition and associative facilitation as affected by degree of prior learning. *J. exp. Psychol.*, 1953, **46**, 400–404.

Ausubel, D., Robbins, L., & Blake, E., Jr. Retroactive inhibition and facilitation in the learning of school materials. *J. educ. Psychol.*, 1957, **48**, 334–343.

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Belbin, E. The influence of interpolated recall on recognition. *Quart. J. exp. Psychol.*, 1950, **2**, 163–169.

Bilodeau, I. M., & Schlosberg, H. Similarity in stimulating conditions as a variable in retroactive inhibition. *J. exp. Psychol.*, 1951, **41**, 199–204.

Blankenship, A. B., & Whitely, P. L. Proactive inhibition in the recall of advertising material. *J. soc. Psychol.*, 1941, **13**, 311–322.

Boring, E. G. Communality in relation to proaction and retroaction. *Amer. J. Psychol.*, 1941, **54**, 280–283.

Briggs, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

Briggs, G. E. Retroactive inhibition as a function of degree of original and interpolated learning. *J. exp. Psychol.*, 1957, **53**, 60–67.

Briggs, G. E., Thompson, R. F., & Brogden, W. J. Retention functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **48**, 419–423.

Bugelski, B. R. Interference with recall of original responses after learning new responses to old stimuli. *J. exp. Psychol.*, 1942, **30**, 368–379.

Bugelski, B. R. An attempt to reconcile unlearning and reproductive inhibition explanations of proactive inhibition. *J. exp. Psychol.*, 1948, **38**, 670–682.

Bugelski, B. R., & Cadwallader, T. A reappraisal of the transfer and retroaction surface. *J. exp. Psychol.*, 1956, **52**, 360–366.

Buxton, C. E. List structure as a determiner of amount of retroactive inhibition. *Psychol. Bull.*, 1941, **38**, 719. (Abstract)

Cassel, R. H. Serial verbal learning and retroactive inhibition in aments and normal children. Unpublished doctoral dissertation, Northwestern Univer., 1957.

Ceraso, J. An experimental critique of competition of response and specific interference as factors in retroactive inhibition. Unpublished doctoral dissertaion, New School for Social Research, 1959.

Cooper, J. B. An attempt to measure the "tension" values of interpolated situations. *J. gen. Psychol.*, 1942, **27**, 347–351.

Crannell, C. W. An effective demonstration of retroactive and proactive inhibition. *Amer. J. Psychol.*, 1948, **61**, 391–395.

Davis, D. R., & Sinha, D. The effect of one experience on the recall of another. *Quart. J. exp. Psychol.*, 1950, **2**, 43–52. (a)

Davis, D. R., & Sinha, D. The influence of an interpolated experience upon recognition. *Quart. J. exp. Psychol.*, 1950, **2**, 132–137. (b)

Deese, J., & Hardman, G. W. An analysis of errors in retroactive inhibition of rote verbal learning. *Amer. J. Psychol.*, 1954, **67**, 299–307.

Deese, J., & Marder, V. J. The pattern of errors in delayed recall of serial learning after interpolation. *Amer. J. Psychol.*, 1957, **70**, 594–599.

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

Gibson, E. J. Retroactive inhibition as a

function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.

Gladis, M., & Braun, H. W. Age differences in transfer and retroaction as a function of intertask response similarity. *J. exp. Psychol.*, 1958, **55**, 25–30.

Greenberg, R., & Underwood, B. J. Retention as a function of stage of practice. *J. exp. Psychol.*, 1950, **40**, 452–547.

Greenspoon, J., & Ranyard, R. Stimulus conditions and retroactive inhibition. *J. exp. Psychol.*, 1957, **53**, 55–59.

Haagen, C. H. Learning and retention as a function of the synonymity of original and interpolated tasks. Unpublished doctoral dissertation, State Univer. Iowa, 1943.

Haagen, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **30**, 185–200.

Hall, J. F. Retroactive inhibition in meaningful material. *J. educ. Psychol.*, 1955, **46**, 47–52.

Hamilton, R. J. Retroactive facilitation as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1943, **32**, 363–376.

Heise, G. Retroactive inhibition as a function of the degree of approximation to English word order. *Amer. Psychologist*, 1956, **11**, 450. (Abstract)

Highland, R. W. Retroactive inhibition: Effects of S-R variations in relation to degree of interpolated learning. Unpublished doctoral dissertation, Ohio State Univer., 1949.

Houlahan, F. J. Immediacy of interpolation and amount of inhibition. *J. educ. Psychol.*, 1941, **32**, 37–44.

Irion, A. L. Retroactive inhibition as a function of the relative serial positions of the original and interpolated items. *J. exp. Psychol.*, 1946, **36**, 262–270.

Irion, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **55**, 336–341.

Ishihara, I. The process of retroactive inhibition in retention. *Jap. J. Psychol.*, 1951, **21**, 18–25.

Ishihara, I., & Kasha, K. The learning of response words in similar, opposite, or neutral relation: A study on the conditioning principle in verbal learning. *Jap. J. Psychol.*, 1953, **24**, 1–12.

Ishihara, I., Morimoto, H., Kasha, K., & Kubo, K. Associative directions and semantic relations in verbal learning. *Tohoku psychol. Folia*, 1957, **16**, 7–18.

Jenkins, W. O., & Postman, L. An experimental analysis of set in rote learning: Retroactive inhibition as a function of changing set. *J. exp. Psychol.*, 1949, **39**, 69–72.

Jones, W. F., Jr. A comparison of retroactive and proactive inhibition as a function of the time interval between original learning and the measurement of retention. Unpublished doctoral dissertation, Vanderbilt Univer., 1953.

Kennelly, T. W. The role of similarity in retroactive inhibition. *Arch. Psychol.*, 1941, **37**, No. 260.

King, D. J., & Cofer, C. N. Retroactive interference in meaningful material as a function of the degree of contextual constraint in the original and interpolated learning. *ONR tech. Rep.*, 1958, No. 21. (Contract NONR 595 (04), Univer. Maryland)

Kingsley, H. L. The factors of similarity and association in retroactive inhibition. *Amer. Psychologist*, 1946, **1**, 262. (Abstract)

Livson, N. H., & Krech, D. Dynamic systems, rote learning, and retroactive inhibition. *J. Pers.*, 1955, **24**, 2–19.

McClelland, D. C., & Heath, R. M. Retroactive inhibition as a function of degree of association of original and interpolated activities. *J. exp. Psychol.*, 1943, **33**, 420–430.

McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

McGeoch, J. A., & Underwood, B. J. Tests of the two-factor theory of retroactive inhibition. *J. exp. Psychol.*, 1943, **32**, 1–16.

McMullin, T. E. A study of the affective nature of the interpolated activity as a factor in producing differing relative amounts of retroactive inhibition in recall and in recognition. *J. exp. Psychol.*, 1942, **30**, 201–215.

Maeda, Y. Zur experimentellen Untersuchung über Faktoren der Reproduktionshemming: I. Über Hemmungswirkungen auf die Reproducktion. *Jap. J. Psychol.*, 1951, **21**, 1–17.

Melton, A. W. Overt interlist intrusions and retroactive inhibition as a function of the ratio of the degrees of learning of original and interpolated verbal habits. *Psychol. Bull.*, 1941, **38**, 575.

Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.

Melton, A. W., & von Lackum, W. J. Retroactive and proactive inhibition in retention: Evidence for a 2-factor theory of

retroactive inhibition. *Amer. J. Psychol.*, 1941, **54**, 157–173.

MILLER, G. A. *Language and communication.* New York: McGraw-Hill, 1951.

MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.

MORGAN, R. L., & UNDERWOOD, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40**, 592–603.

MORROW, M. A. The relation of overt errors during learning to transfer and retroactive inhibition. Unpublished doctoral dissertation, Washington Univer., 1954.

NEWTON, J. M. Interlist similarities and point of interpolation in retroactive inhibition of verbal associations. Unpublished doctoral disertation, Ohio State Univer., 1955.

NEWTON, J. M., & WICKENS, D. D. Retroactive inhibition as a function of the temporal position of the interpolated learning. *J. exp. Psychol.*, 1956, **51**, 149–154.

OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 277–301.

OSGOOD, C. E. An investigation into the causes of retroactive interference. *J. exp. Psychol.*, 1948, **38**, 132–154.

OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.

PARDUCCI, A., & KNOPF, N. B. Retroactive facilitation when new responses have been learned to old stimuli. *Amer. J. Psychol.*, 1958, **71**, 426–428.

PEAIRS, R. H. Development and analysis of retroactive inhibition in retention of meaningful connected verbal stimulus material. Unpublished doctoral dissertation, Ohio State Univer., 1958.

PEIXOTTO, H. E. Proactive inhibition in the recognition of nonsense syllables. *J. exp. Psychol.*, 1947, **37**, 81–91.

PETERSON, L. R., & PETERSON, M. J. Intrusions at recall in retroactive inhibition. *Amer. Psychologist*, 1957, **12**, 419.

POSTMAN, L. Retroctive inhibition in recall and recognition. *J. exp. Psychol.*, 1952, **44**, 165–169.

POSTMAN, L. Mediated equivalence of stimuli and retroactive inhibition. *Amer. J. Psychol.*, 1958, **71**, 175–185.

POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: III. Interserial interference. *J. exp. Psychol.*, 1956, **51**, 323–328.

POSTMAN, L., & ALPER, T. G. Retroactive inhibition as a function of the time of interpolation of the inhibition between learning and recall. *Amer. J. Psychol.*, 1946, **59**, 439–449.

POSTMAN, L., EGAN, J. P., & DAVIS, J. Rate of recall as a measure of learning: I. The effects of retroactive inhibition. *J. exp. Psychol.*, 1948, **38**, 535–546.

POSTMAN, L., & KAPLAN, H. L. Reaction time as a measure of retroactive inhibition. *J. exp. Psychol.*, 1947, **37**, 136–145.

POSTMAN, L., & POSTMAN, D. L. Change in set as a determinant of retroactive inhibition. *Amer. J. Psychol.*, 1948, **61**, 236–242.

POSTMAN, L., & RILEY, D. A. Degree of learning and interserial interference in retention. *U. Calif. Publ. Psychol.*, 1959.

PRENTICE, W. C. H. Retroactive inhibition and the motivation of learning. *Amer. J. Psychol.*, 1943, **56**, 283–292.

RAY, W. S. Proactive inhibition: A function of time interval. *Amer. J. Psychol.*, 1945, **58**, 519–529.

RICHARDSON, J. Retention of concepts as a function of degree of original and interpolated learning. *J. exp. Psychol.*, 1956, **51**, 358–364.

RITCHIE, M. L. The Skaggs-Robinson hypothesis as an artifact of response definition. *Psycho. Rev.*, 1954, **61**, 267–270.

ROTHKOPF, E. Z. A deduction from an excitation-inhibition account of retroactive inhibition. *J. exp. Psychol.*, 1957, **53**, 207–213.

RUNQUIST, W. N. Retention of verbal associates as a function of strength. *J. exp. Psychol.*, 1957, **54**, 369–375.

SALTZ, E. Act of regression as a special case of retroactive inhibition and functionally related to stimulus generalization. *J. exp. Psychol.*, 1953, **45**, 394–400.

SEELER, R. A. Acquisition and retention of verbal materials. Unpublished honors thesis, Univer. Vermont, 1958.

SEIDEL, R. J. The concurrent effects of proactive and retroactive inhibition. *J. exp. Psychol.*, 1959, **57**, 397–402.

SHAKLEE, A. B., & JONES, B. E. Problems of method and theory in controlling rest activity. *J. gen. Psychol.*, 1959, **60**, 11–16.

SHAW, F. J. Influence of degree of original learning upon associative and reproductive inhibition. *Proc. Ia. Acad. Sci.*, 1942, **49**, 413–417.

SHERMAN, L. J. Retention in psychopathic, neurotic, and normal subjects. *J. Pers.*, 1957, **25**, 721–729.

SHRIVER, E. L. An artifact of retroactive inhibition. *Amer. Psychologist*, 1953, **8**, 435. (Abstract)

SLAMECKA, N. J. Studies of retention of con-

nected discourse. *Amer. J. Psychol.*, 1959, **72**, 409–416.

Slamecka, N. J. Retroactive inhibition of connected discourse as a function of practice level. *J. exp. Psychol.*, 1960, **59**, 104–108. (a)

Slamecka, N. J. Retroactive inhibition of connected discourse as a function of similarity of topic. *J. exp. Psychol.*, 1960, **60**, 245–249. (b)

Swenson, E. J. Retroactive inhibition: A review of the literature. *Minn. Stud. Educ.*, 1941, No. 1.

Thune, L. E. Reproductive interference following appropriate and inappropriate warm-up activities. *J. exp. Psychol.*, 1958, **55**, 535–542.

Thune, L. E., & Underwood, B. J. Retroactive inhibition as a function of degree of interpolated learning. *J. exp. Psychol.*, 1943, **32**, 185–200.

Twedt, H. M., & Underwood, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, **58**, 111–116.

Underwood, B. J. The effects of punishment in serial verbal learning. *Proc. Ia. Acad. Sci.*, 1941, **48**, 349–352.

Underwood, B. J. A test of the two-factor theory of retroactive inhibition by use of the paired-association technique. *Psychol. Bull.*, 1942, **39**, 593. (Abstract)

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**(3, Whole No. 273).

Underwood, B. J. Retroactive and proactive inhibition after 5 and 48 hours. *J. exp. Psychol.*, 1948, **38**, 29–38. (a)

Underwood, B. J. "Spontaneous recovery" of verbal associations. *J. exp. Psychol.*, 1948, **38**, 429–439. (b)

Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949. (a)

Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34. (b)

Underwood, B. J. Proactive inhibition with increased recall time. *Amer. J. Psychol.*, 1950, **63**, 594–599. (a)

Underwood, B. J. Retroactive inhibition with increased recall time. *Amer. J. Psychol.*, 1950, **63**, 67–77. (b)

Underwood, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.

Waters, R. H. The concept of psychological disposition and retroactive inhibition. *Psychol. Bull.*, 1941, **38**, 573. (Abstract)

Waters, R. H. Degree of learning and proactive inhibition in retention. *Psychol. Bull.*, 1942, **39**, 495–496. (Abstract)

Werner, H. The effect of boundary strength on interference and retention. *Amer. J. Psychol.*, 1947, **60**, 598–607.

Wywrocki, E. H. Age difference in retroactive inhibition as a function of the degree of similarity of serial position between the original and interpolated learning. Unpublished doctoral dissertation, Univer. Pittsburgh, 1957.

Young, R. K. Retroactive inhibition and proactive inhibition under varying conditions of response similarity. *J. exp. Psychol.*, 1955, **50**, 113–119.

(Received June 20, 1959)

Psychological Bulletin
1960, Vol. 57, No. 6, 476–492

A SEARCH FOR RELATIONS BETWEEN BRAIN CHEMISTRY AND BEHAVIOR[1]

MARK R. ROSENZWEIG, DAVID KRECH, AND EDWARD L. BENNETT
University of California, Berkeley

In 1953 we embarked on an experimental program to test the general hypothesis that variation in brain chemistry is a major determinant of the variation in adaptive behavior among normal individuals. The experimental animal was the rat, the biochemical measure first employed was cholinesterase (ChE) activity, and the adaptive behavior initially studied was "hypothesis preference" as previously defined by Krechevsky (1932).

This program was initiated with a specific experimental hypothesis to be tested. Our first completed experiment indicated that this hypothesis was not tenable, but it did suggest that a revision of the original hypothesis could accommodate the data. This revision (our second experimental hypothesis) was then tested in a series of experiments, some of whose results supported the second hypothesis, while others were more ambiguous in their implications. A long-term genetic experiment had meanwhile been started with the hope that it would provide crucial information concerning the tenability of our second hypothesis. This experiment, when completed, indicated that our second hypothesis needed revision. This then led to the third formulation. This third experimental hypothesis was then subjected to further tests—and now all the tests made so far seem to confirm this hypothesis.

Throughout the vicissitudes outlined above, our most general hypothesis has been abundantly justified by the results accumulated—significant relationships have been demonstrated to exist between variations in brain biochemistry and variations in adaptive behavior.[2] This will be shown here by a review of the experimental results obtained in the course of our research, some of which have been previously published and some of which have not. In this paper we will trace the course of our theorizing and experimentation. A systematic account of the hypotheses left behind us, as well as those which continue to guide us, may help to make clear the relations of our successive research reports which have appeared in a variety of journals.

First Experimental Hypothesis

Our initial experimental hypothesis related specific behavior preferences to ChE activity in defined cortical areas. The test was conducted in the Krech Hypothesis Apparatus. This apparatus includes four successive choice points. At each one, the rat could enter either the left or the right alley. One

[1] This investigation was supported in part by a research grant, M-1292, from the National Institute of Health, United States Public Health Service, and in part by a research grant, G-2542, from the National Science Foundation; it also received support from the United States Atomic Energy Foundation. During 1958–59 the two first authors were Research Professors in the Institute for Basic Research in Science, University of California. We also want to acknowledge our debt to M. Calvin who has been interested in this study since its inception and who has helped our work in many ways.

[2] While we have found that brain chemistry and behavior are interdependent, we will uot consider in this paper our studies in which brain biochemistry is treated as the *dependent* variable, being affected by training (Krech, Rosenzweig, & Bennett, in press) or by brain lesions (Krech, Rosenzweig, & Bennett, 1960).

of the alleys was illuminated and the other was dark; at half of the choice points the light was on the left side, and at half it was on the right side. One of the two alleys was open at its far end, permitting the animal to progress toward the goal; the other alley was blocked, forcing the animal to retrace. After each run, the pattern of lighted alleys and open alleys was changed in a prescribed order so that the problem was unsolvable, since neither the left or the right alleys (spatial cues) nor the lighted or dark alleys (visual cues) led consistently to the goal. Confronted with this problem, most animals, nevertheless, followed one of the cues rather consistently on any one day. The rat might consistently choose the right alleys or the left alleys (spatial preferences) or it might consistently choose the lighted or dark alleys (visual preferences). The ChE determinations for each animal were made separately on samples of tissue removed from three areas of the dorsal cortex —the visual, somesthetic, and motor areas. The values obtained were measures of the *activity* of the enzyme in breaking down acetylcholine under standard conditions. The methods used are detailed in Krech, Rosenzweig, and Bennett (1956) and Rosenzweig, Krech, and Bennett (1958b).

The specificity hypothesis was stated as follows: *the rat's spatial or visual preferences in attempting to solve a spatial-visual problem are a function of the relatively greater ChE activity in the somesthetic or visual area, respectively, of the cerebral cortex.* This hypothesis was based on the earlier observation that spatial or visual preferences can be systematically obtained by making lesions in the appropriate cortical regions, somesthetic or visual (Krechevsky, 1935), and by the assumption that ChE activity in any cortical region was an index of the transmission efficiency of that region.

The initial study, reported in a brief communication to *Science* (Krech, Rosenzweig, Bennett, & Krueckel, 1954), however, showed that rats with strong spatial preferences had significantly greater ChE activity than rats with strong visual preferences not only in the somesthetic but also in the visual and motor areas of the cortex. In other words, the spatial rats had greater ChE activity *throughout* the dorsal cortex, compared with the visual rats. These data were clearly inconsistent with our expectations, and we therefore abandoned the first experimental hypothesis. But if the data were compellingly inconsistent with our specificity hypothesis, they were equally compelling in suggesting an alternative hypothesis—one which would relate the animal's behavior to its *general* level of cortical ChE activity.

Second Experimental Hypothesis

Our second hypothesis, which might be termed a generality hypothesis, rested on the assumption that animals with spatial preferences in our tests were generally superior in adaptive behavior to animals with visual preferences. With this assumption, our hypothesis was stated as follows: *overall levels of cortical ChE activity are positively related to efficiency of synaptic transmission and are therefore positively correlated with problem-solving ability.* Of particular interest for this hypothesis was the observation of correlated differences in behavior and brain chemistry among three of the strains of animals used in the experiment: the S_1 strain (descendants of the Tryon maze-bright animals, Tryon, 1940), the S_3 strain (descendants of the Tryon maze-dull animals), and descendants of a cross between those two strains. The S_1 rats, which could be presumed to be the best learners, showed spatial preferences and were relatively high in ChE activity. The S_3 rats, which could be presumed to be the poorest learners, showed visual preferences and were relatively low

in ChE, and the crosses tended to overlap both groups.

While this second hypothesis was outlined in our initial publication (Krech, et al., 1954), it was spelled out more fully—in both its behavioral and its biochemical implications—in the 1955 Wisconsin Symposium (Rosenzweig, Krech, & Bennett, 1958a). The assumption that animals with spatial preferences were generally superior in adaptive behavior to animals with visual preferences was supported by the following observations. We were using the unsolvable problem in the Krech Hypothesis Apparatus; that is, the rat was equally likely to be successful at a choice point whether he chose the light or dark alley or the right or left side. Under standard conditions of training, almost all animals nevertheless show a light-going (visual) preference on their first trials (see Rosenzweig, et al., 1958a, Fig. 117). Since the tests are made in a darkened room, this was interpreted to mean that the animals respond to the most dominant and insistent cue in the problem situation. An animal could therefore achieve a spatial preference only by ignoring the dominant cue, abandoning its initial hypothesis, and adopting the less obvious hypothesis of location in space. The ability of an animal to pay attention and respond to various stimulus aspects of its environment when confronted with a problem was, we assumed, positively correlated with its general efficiency in adaptive behavior. We therefore concluded that the spatial animal was more adaptive than the animal that could not leave off its response to the dominant visual cue and would therefore end with a visual preference score.

On the biochemical side we made specific, in the Wisconsin Symposium, two points that had only been suggested in the initial preliminary report to *Science:* the major biochemical variable upon whose assumed function our hypothesis rested, was the transmitter substance, acetylcholine (ACh); and ChE activity in our hypothesis was employed as an index to the availability of ACh.

Our argument in support of these points was as follows: (*a*) The transmitter substance, ACh, is importantly involved in neural transmission in the central nervous system. It is now generally acknowledged that when a neural impulse reaches the end of a presynaptic neuron, a chemical transmitter is released which diffuses across the synapse and excites the post-synaptic neuron. (For a recent review of chemical transmission in the central nervous system, see Crossland, 1960.) ACh is the transmitter substance at many peripheral and central synapses. The released ACh, after exciting the post-synaptic neuron, is promptly hydrolyzed and inactivated by the enzyme ChE, and the synaptic junction is returned to its prior state. (*b*) Animals with greater availability of ACh could therefore show readier transmission of nerve impulses. (*c*) Animals with relatively efficient transmission systems (higher rates of ACh functioning) would tend to manage new problems more effectively than animals with less efficient transmission systems (low ACh functioning rates). This latter assumption was tempered by the phrase "within limits" since it is possible that beyond some point, ready transmission of impulses could result in explosive, undisciplined, and unintegrated behavior. (*d*) The ACh transmitter system includes two enzymes, choline acetylase (ChA) which synthesizes ACh, and ChE which breaks down ACh. While our major

concern was with ACh, we decided to use one of the enzymes as an index rather than to measure ACh concentration directly, for the following reasons:

ACh is synthesized and destroyed continuously, and its concentration alters rapidly even with temporary changes in the functional state of the animal. Furthermore, when an animal is sacrified for assay, special precautions must be employed to stop at once all chemical activity in the brain, otherwise a large part of ACh will quickly be destroyed by the ChE present. These precautions involve dealing with frozen brain, from which it is difficult to take precisely defined cortical samples. Finally, no reliable chemical technique is at present available for the measurement of ACh; only bio-assay methods can be used. Of the ChA-ACh-ChE system, ChE is a stable component for whose measurement we had available a relatively simple and extremely reliable chemical technique (Neilands & Cannon, 1955). This fact, together with some evidence which indicated that ChE was related to the activity levels of the other members of the system led to the decision to use ChE as an index to the availability of its substrate, ACh. However, we made this decision with some misgivings and noted that it would be important "to test the use of ChE as an index to ACh metabolism, since this is not founded upon direct observation" (Rosenzweig, et al., 1958a, pp. 397–398).

Thus, our second experimental hypothesis—that a positive correlation exists between efficiency in adaptive behavior and the overall ChE level of the cortex—rested upon three major assumptions: animals which showed a spatial preference in the Krech Hypothesis Apparatus had a higher adaptive capacity than animals which showed a visual preference, the availability of cortical ACh was positively related to synaptic transmission efficiency, cortical ChE activity provided a good index to the availability of ACh.

Supporting Data

Three types of experimental evidence appeared to offer support for the second experimental hypothesis in our 1955 report: (*a*) Tests of additional animals continued to show that those with spatial preferences had greater cortical ChE activity than those with visual preferences (Rosenzweig, et al., 1958a, pp. 385–388).[3] (*b*) The intercorrelations among ChE measures in the three cortical areas (visual, somesthetic, and motor) were consistently positive and of the order of .60 (Rosenzweig, et al., 1958a, Table 30.) This supported the suggestion that the level of ChE activity of the rat's cortex is a general characteristic of its brain biochemistry. (*c*) Small doses of pentobarbital sodium were found to fixate the animals' initial preferences and to prevent them from shifting to spatial hypotheses (Rosenzweig, Krech, & Bennett, 1956, 1958a, pp. 391–396). The use of this drug was suggested by the fact that in appropriate small concentrations it has been shown *in vitro* to reduce the rate of synthesis of ACh (McLennan & Elliott, 1951), and we therefore expected it to lower adaptive capacity. The behavioral effect was in accordance with our prediction, and Moroz (1959) found that pentobarbital could also fixate animals on spatial preferences. Nevertheless, we do not now know how much weight should be given to this evidence, for the following reasons: we did not determine whether the rate of synthesis of ACh was actually reduced in the brains of our animals, and extrapolation from the *in vitro* to the *in vivo* condition is problematical; pentobarbital has other effects in

[3] Pierce (1959) used a problemless version of this test (the blocking doors having been removed) in which the behavior presumably afforded a measure of sensory preference; in this case the correlation of preference with ChE activity was significant and opposite in sign from what we had obtained in the problem-solving situation.

the brain which also tend to reduce the brain's responsiveness.

In 1956 (Krech, et al.), a new test of the hypothesis was employed. Instead of presenting the animal with an unsolvable problem—the standard procedure for testing for hypothesis preference—a "progressively solvable" problem in the Krech Hypothesis Apparatus was employed. On the first day of testing the animals were presented with the usual unsolvable problem, during which they tended to adopt a visual hypothesis. The animals were then divided into two groups: for one group, the left alley was made progressively more often correct on succeeding days (the spatial problem); for the other group, the lighted alley was made progressively more often correct on succeeding days (the visual problem). The results of this experiment with a partially solvable problem were consistent with our prediction that high ChE activity favored relatively rapid shifts in the animal's hypotheses when the situation demanded such shifts. This generalization held for both problems.

The next step in the testing of our hypothesis was to determine whether relations exist between *learning* and ChE activity, using standard animal learning tasks. We were encouraged to take this step, having already found a relationship between *hypothesis preferences* and ChE in an *unsolvable problem* and between *hypothesis shifts* and ChE in a *partially solvable* problem, using the Krech Hypothesis Apparatus for both problems. Furthermore, we wanted to follow-up the observation concerning strain differences made at the very start of our research—that the S_1 strain showed higher ChE activity than the S_3 strain. This observation had meanwhile been repeatedly confirmed by the chemical analysis of large numbers of S_1 and S_3 rats run in succeeding experiments. In these two strains then, we had animals of known ChE characteristics. We therefore attempted to determine whether the animals of the S_1 strain would be superior to those of the S_3 strain in their performance on the various learning tasks. Indeed, if our hypothesis had any validity, it was necessary to predict such strain differences in behavior.

Our first exploratory experiment (in 1956, unpublished) tested animals of these two strains on the Lashley III Maze. The results, although based on rather small groups, were encouraging. We then began a more systematic examination of the learning performance of these two strains, using three different learning tasks: the Lashley III Maze, the Hebb-Williams Maze, and the Dashiell Maze. A total of 77 S_1 and 80 S_3 male rats was divided among the three tests. In each of these three tests the S_1 animals made significantly fewer errors than the S_3 animals. (See Table 1—the data for 139 of these S_1 and S_3 rats were reported at the 1958 APA meeting—Rosenzweig, Krech, Bennett, & Longueil, 1958c—and the data for all 157 animals were presented at the 1959 Pittsburgh Symposium—Rosenzweig, Krech, & Bennett, 1959.)

About this time we obtained other biochemical measures which seemed to offer further support for our hypothesis. The observed relationship between ChE and learning ability was assumed, of course, to reflect the fact that the ACh–ChE system is of peculiar significance to cerebral functioning because of its contribution to synaptic transmission. However, since the ACh–ChE system is only one of a number of biochemical sys-

TABLE 1

STRAIN DIFFERENCES IN BEHAVIORAL TESTS

Maze	Mean Errors		Correlations: Errors *vs.* ChE		Mean Errors			
	S_1	S_3	Signs of indiv. *r*'s	Combined *r*	RCH	RCL	RDH	RDL
Hebb-Williams								
Score	66***	78	−++	.03	90	86	90*	82
N	(37)	(39)		(30)	(20)	(20)	(26)	(30)
Dashiell								
Score	24***	33	++	.52**	28***	18	18	20
N	(28)	(27)		(24)	(25)	(25)	(30)	(30)
Lashley III								
Score	19***	42	+++	.18	34**	24	24	24
N	(14)	(14)		(33)	(22)	(22)	(22)	(22)

Note.—Most of these data were presented at the 1959 Pittsburgh Symposium (Rosenzweig, et al., 1959), but data for 93 additional animals of the Roderick strains are included here (25 additional RCH, 25 RCL, 21 RDH, and 22 RDL).

* $p < .10$.
** $p < .05$.
*** $p < .01$.

tems which are important in brain functioning, the possibility remained that ChE activity merely reflected the *general* enzyme and metabolic levels in the rat brain. The observed behavior-ChE relations would not, in this case, depend upon the specific role of ChE in synaptic transmission. We therefore started on a program to determine what relationships exist between measures of ChE and other substances. Our first experiment (Bennett, Krech, Rosenzweig, Karlsson, Dye, & Ohlander, 1958a) in this series compared ChE activity with lactic dehydrogenase (LDH) activity in the cerebral cortex and subcortex of S_1 and S_3 rats. (The term "subcortex" was used to refer to the whole brain after the dorsal cortex had been removed.) LDH was chosen since it is an enzyme important in brain metabolism but apparently plays no specific role in the transmission process.

The results, based on ChE and LDH analyses of 106 male animals of several age groups, showed clear differences between distribution of activity of the two enzymes, for example: ChE activity showed a more sharply differentiated pattern of regional distribution within the cortex than did LDH; within the age range studied (30 to 150 days), the change in enzymic activity with increasing age was more marked for ChE than for LDH; within each strain separately, there was some indication of a modest positive correlation between cortical LDH and ChE—for the subcortex, there was no correlation; where the S_1 animals were, of course, higher in ChE activity than the S_3 animals for both cortex and subcortex, *no differences were found between the strains for LDH.*

The second study in this series (Bennett, Rosenzweig, Krech, Ohlander, & Morimoto, in press) examined the relationship between ChE and percent brain protein. In general we found the same results that

we had in our LDH study. This time, however, we had available animals from four Roderick strains (to be described in the next section), as well as the S_1 and S_3 strains. The brains of 226 animals, ranging in age from 6 to 470 days, were chemically analyzed for percent protein and ChE activity. Some of the main differences found for all strains between ChE and percent protein were the following: relative values for different regions of the cortex differed greatly for the two substances; ChE activity was found to be much more dependent upon age than was percent protein; ChE and percent protein showed no correlation with each other in cortex, and only a moderately low correlation in subcortex; *pairs of strains differing markedly in ChE activity did not differ in percent protein.*

The results of both these studies ruled out the interpretation that correlations were obtained between ChE activity and behavior only because ChE activity simply reflects general biochemical characteristics of the brain. The lack of correlation between ChE activity and other measures of biochemical activity in the brain is consistent with our hypothesis that the observed relations between behavior and ChE activity are to be attributed to the specific role of ChE activity in synaptic transmission.

At the 1955 Wisconsin Symposium the criticism was made that the differences in ChE activity that we had reported among our various experimental animals were so small as to be "close to or within the probable limits of, error of sampling and analysis" (Tower, 1958, p. 356). It is true that we are dealing with small differences, but anyone working with brain biochemistry must be prepared to deal with small differences; we have repeatedly shown that the biochemical variability of the brain is small compared with that of other organs (Bennett, et al., 1958a; Bennett, Rosenzweig, Krech, Karlsson, Dye, & Ohlander, 1958b; Bennett, et al., in press). This necessitates having reliable and valid measures of brain ChE. Two kinds of evidence make it clear that the small differences we find represent true individual and strain differences that cannot be attributed to "error of sampling and analysis": (*a*) In a paper entitled "Individual, Strain and Age Differences in Cholinesterase Activity of the Rat Brain" (Bennett, et al., 1958b), we demonstrated, using data from over 400 animals, that the observed differences between strains and among ages were consistent and highly reliable. (*b*) A successful genetic selection experiment, to be described later in this paper, was based solely on measurements of cortical ChE activity; if the observed differences represented only errors, no effect of selection on ChE activity could have been achieved.

Our original hypothesis was now bolstered by the observation of consistently positive correlations between hypothesis behavior and cortical ChE among individuals, and between learning performance and cortical ChE among strains, and by evidence that variations in ChE activity did not merely reflect variations in general biochemical characteristics of the brain. But these behavioral-biochemical correlations were not completely satisfactory supports for our hypothesis. The hypothesis asserted that behavior differences are caused by, or are a function of biochemical differences. Obviously it is hazardous to argue from correlations to existence of a

causal relation. It was necessary, in other words, to devise an experiment which would answer the following question: were the observed correlations between our behavioral and biochemical variables *necessary* or were they fortuitous?

The traditional procedure in seeking to answer such a question would be to manipulate experimentally the biochemical variable (the independent variable) and observe whether learning performance (the dependent variable) changed in an appropriate manner. Thus, for example, one could alter the rate of ACh synthesis or the level of ChE activity by the use of drugs and see whether predicted changes would occur in behavior. We had used such an approach when we employed pentobarbital sodium to inhibit ACh synthesis and predicted that the injected animals would show less adaptive problem-solving behavior than control animals (Moroz, 1959; Rosenzweig, et al., 1956). The behavioral results had been consistent with our assumptions about the role of ACh functioning, but, as we have indicated, these results cannot be interpreted unambiguously. In another attempt along this line in our laboratory, McGaugh and Petrinovich (1959) found that small doses of strychnine improved the rate of learning in rats. There is evidence that strychnine may inhibit ChE activity (Nachmansohn, 1938), but this drug, too, has other actions, so that its effects here cannot be attributed to the ACh–ChE system. Successful use of a drug which is more specific in its effect on ChE activity had been indicated by Russell (1954) in a brief abstract entitled, "Effects of Reduced Brain Cholinesterase on Behavior." We also tried using a specific ChE-inhibitor, diisopropylfluorophosphonate (DFP), but we were finally led to abandon it and have not previously reported on these attempts. Dosages of DFP large enough to produce substantial reductions in brain ChE had deleterious effects on gastrointestinal functioning. We therefore feared that the animals' motivational state might not be comparable with that of the control animals, since we used food deprivation as the motivating condition in our behavioral testing. Furthermore, DFP was lethal for some members of each experimental group we employed (despite comparable dosages by body weight), and there was no way of knowing how this selective action of the drug would bias the results. The animals which did survive the drug and completed the tests showed no consistent differences in performance from their controls.

Genetic Experiments

Meanwhile we were preparing to use another approach to the problem. Genetic selection experiments, it occurred to us, offered a possible way of developing strains with relatively high or low ChE activity. These strains could then be tested for behavioral differences. In 1955, Roderick, in our laboratory, under the guidance of E. R. Dempster of the Department of Genetics, began a selective breeding program designed to create strains of animals differing in cortical ChE activity. Roderick started his selective breeding from two heterogeneous foundation stocks from the Genetics Laboratory—the Castle and the Dempster stocks. From each of these foundation stocks Roderick bred a high- and a low-ChE line. This program required three years (six selected generations) for its successful completion and at the

end we had available four strains of animals of known ChE activity; the RDH (Roderick-Dempster High ChE) and the RDL strains, the RCH and RCL strains (Roderick, in press).

With these strains available to us we would be in a position to make a crucial test of the significance of our previously observed correlations between behavior and cortical ChE. If our hypothesis was valid, we should find the RDH strain was superior to the RDL strain, and the RCH strain superior to the RCL strain. On the other hand, if there were no necessary relation between learning ability and ChE activity, then no consistent behavioral differences would be found between the RDH and RDL strains or between the RCH and RCL strains.

While this selective breeding program was being carried out, a second genetic approach was also begun. This approach involved the use of a strain prepared by crossing the S_1 and S_3 strains. The cross, called the K strain, was begun in 1957 and was then interbred randomly for several generations. There were two considerations which led to the creation of the K strain.

1. If our hypothesis was wrong and it was merely fortuitous that the S_1 strain was both superior in learning ability and higher in cerebral ChE activity than the S_3 strain, then the association between the behavioral and biochemical traits would be expected to disappear among the animals of the K strain. That is, correlations between error scores and ChE activity within the K strain would tend to approach zero if the genetic determinants of learning ability and brain ChE reassorted randomly. If, on the other hand, our hypothesis was valid and there was an intrinsic relation between cerebral ChE activity and learning ability, those animals of the K strain that had higher ChE activity would tend to show fewer errors in the learning tasks; that is, correlations between ChE and errors would be negative.

2. The second consideration which suggested the desirability of working with an $S_1 \times S_3$ cross derived from the same reasoning. Our previous tests involving correlational analysis were hampered by the restricted range of ChE activity within the S_1 and S_3 strains. It was thought that the K strain would be more heterogeneous than the S_1 or the S_3 strain, and that using the K strain would therefore increase the probability of obtaining a significant correlation, if an intrinsic correlation between behavior and ChE did in fact exist.

Before the K strain was ready for testing, some unexpected (and premonitory) results were obtained by McGaugh (1959) in our laboratories. He used descendants of crosses (the S_{13} and M_2 strains) established a number of years ago between the S_1 and the S_3 strains. When these animals were tested on a 14-unit alley maze, McGaugh found a significant *positive* correlation between cortical ChE activity and errors ($n = 26$, $r = .47$, $p < .05$). On the Lashley III Maze, he found a correlation that was not significant, but that was also positive ($n = 13$, $\rho = .28$, $p > .30$). McGaugh found, in other words, some indication that with the crossed strains the higher the ChE activity, the *worse* was the animal's learning performance.

Our results with the K strain (obtained soon after McGaugh's work) are shown in the center columns of Table 1. The animals were tested in eight subgroups. In seven of the eight, positive correlations between ChE activity and errors were found. Thus we had replicated and extended the generality of McGaugh's observation to other learning tasks. For each test, the correlations of the subgroups were combined, using Fisher's r to z transformation. The sizes of the combined correlations tend to be small, and only that for the Dashiell Maze is significant. The consistently *positive* direction of the correlations

was puzzling. It will be remembered that we had been prepared for one of two findings: If our second hypothesis was valid, we should have found *negative* correlations between errors and ChE activity—the higher the ChE activity, the fewer the errors. If, on the other hand, our hypothesis was wrong and no intrinsic relationship existed between ChE activity and learning ability, we should have found *low* and *inconsistent* correlations between the two variables. But our data, as well as those of McGaugh, did not follow either expectation. Whatever the final interpretation of these data would turn out to be, they cast serious doubt on the validity of our second hypothesis.

Before considering the theoretical position further, however, let us look at the results obtained when the Roderick high- and low-ChE strains were tested behaviorally. These results are shown in the right half of Table 1. Following our second hypothesis, we had predicted that the high-ChE strain of each pair would be superior in learning ability to the low-ChE strain derived from the same foundation stock. However, as can be seen from Table 1, in only one of the six comparisons was there any indication that this prediction held (RDH better than RDL on the Dashiell Maze). And this difference was negligible and statistically insignificant. On the other hand, in four cases the low-ChE strain made fewer errors than the high-ChE strain, and three of these differences were significant at the .10 level or better (using a two-tailed test).

We were now faced with a paradox. For the S_1 and S_3 strains it appeared that the higher the ChE activity, the *better* was the learning; for the $S_1 \times S_3$ cross strains (S_{13}, M_2, and K) and for the Roderick strains the higher the ChE activity, the *worse* was the learning. Clearly our second hypothesis, which demanded a positive correlation between ChE activity and learning ability, could not be maintained. Just as clearly, it seemed to us, ChE *was* implicated in learning ability—but the nature of the relationship between ChE and learning ability seemed to differ among strains!

Third Hypothesis

Before all of these results were in, we, together with McGaugh, had begun to explore various revisions of our second hypothesis which would be able to encompass both the differences between the S_1 and S_3 strains and the results with the other strains. In considering the possible mechanism which might lie behind varying relations of ChE activity to learning ability, we were led to re-examine one of the two assumptions basic to our hypothesis—the assumed relation between ChE and ACh. As we have already pointed out, both our first and second hypotheses had rested upon the assumption that ChE activity afforded a good index to the level of ACh functioning. We had accepted this assumption provisionally, pointing out that it should be subjected to a direct test. It then occurred to us that if we revised this basic assumption, a good part of our paradoxical data could be made comprehensible. We therefore made the assumption that ACh and ChE were under relatively *separate* genetic controls, and therefore ChE activity would not be a good index to ACh functioning. Our other basic assumption—that the ACh transmission system, and therefore ChE, was intimately involved in learning ability—was retained. With these two assumptions the present formulation of our third hypothesis reads as

follows: *learning capacity is related to the levels of both ACh and ChE, such that, within limits, the greater the amount of ACh functioning at the synapse, the greater the efficiency of transmission and, consequently, the greater the learning ability.* Greater ACh functioning can be achieved in either of two ways (or by a combination of the two): the more ACh available and released, the greater is the functioning; with a particular amount of ACh released, the *lower* the activity of ChE in breaking down the ACh, the greater is the functioning. This formulation is a revision of several preliminary statements previously made (Krech, Rosenzweig, & Bennett, 1958; McGaugh, 1959; Rosenzweig, et al., 1958c).

This third hypothesis, we believe, can handle the data obtained with the Tryon strains (S_1 and S_3) and with the Roderick and crossed strains.

For the Tryon strains the reasoning is as follows: Tryon's animals that were bred to make few errors in solving a maze would be expected to have nearly optimal ACh functioning. Presumably this would mean relatively high levels of both ACh and ChE, making for sure and rapid synaptic transmission. His animals bred to make many errors would be expected to have unfavorable ACh functioning. Presumably this would mean lower levels of both ACh and ChE, making for less certain and slower transmission.

In the case of the animals of the four Roderick strains which were selected solely for cortical ChE activity, the rate of synthesis and release of ACh might be only slightly affected. Increasing the concentration of ACh has been shown to lead to a compensatory increase in ChE activity (Burkhalter, Jones, & Featherstone, 1957). The converse effect has not been reported; that is, there appears to be no evidence that altering ChE activity will affect the rate of synthesis of ACh. Selection for ChE activity could mean an indirect selection for ACh turnover (despite genetic independence of ChE and ACh), since an individual with greater release of ACh would thereby have a compensatory increase in synthesis of ChE. Nevertheless, since there are other factors controlling the level of ChE activity, genetic selection for ChE should affect ChE more directly and more strongly than it affects ACh. Thus, the high- and low-ChE strains of Roderick might be expected to show little or no difference in ACh. The high-ChE strains might then suffer from too rapid breakdown of ACh at the synapse, making transmission less certain and learning less effective. In the low-ChE strains, released ACh might be allowed to work over a slightly longer period before being broken down, thus rendering synaptic transmission more certain and learning more effective.

For the S_{13}, the M_2, and the K strains we can reason in a similar manner. If ACh metabolism and ChE activity are genetically independent, there would be random reassortment of the genetic determinants for ACh and ChE levels when the Tryon strains are crossed and interbred for several generations. This would mean that any group of descendants of the $S_1 \times S_3$ crosses which showed a high ChE activity would be likely to include animals of high, medium, and low ACh concentration. In other words, on the average, the *high*-ChE animals of the crossed strain would have a *medium* concentration of ACh. In the same way, on the average, *low*-ChE animals of the crossed strain would also have

TABLE 2

ACETYLCHOLINE CONCENTRATIONS AND CHOLINESTERASE ACTIVITY FOR SIX STRAINS

Strain	N	ACh Concentration[a]	N	ChE Activity[b]	ACh/ChE Minutes $\times 10^{2}$[c]
S_1	16	27.3	11	168.3	.162
S_3	19	24.1***	10	153.5**	.157
RCH	11	24.5	11	163.5	.150
RCL	11	22.5*	11	133.4***	.169
RDH	14	23.1	8	146.5	.158
RDL	16	23.0	8	122.7***	.188

[a] ACh concentration is expressed in mμM/gm.
[b] ChE activity is expressed in moles ACh $\times 10^{10}$ hydrolyzed/min/mg.
[c] These values were obtained by dividing the ACh concentration for a strain by its ChE activity, obtained under our assay conditions. This arbitrary value has the dimension of minutes. It expresses the time required for the ACh to be hydrolyzed by the ChE, under our standard conditions. In other words, it would take about one-thousandth of a minute to hydrolyze the ACh in the brain, if this were done under our assay conditions.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

a medium concentration of ACh. From this it follows that, on the average, for the high-ChE animals, the available and released ACh would be too low for optimal ACh functioning, while for the low-ChE animals the released ACh could function over a relatively long period of time. Thus we should expect to find, *on the average* (but with numerous exceptions) that the high-ChE crosses would show less efficient synaptic transmission than the low-ChE and therefore less effective learning. This would result in *low positive* correlations between ChE activity and errors in learning tasks within animals of the $S_1 \times S_3$ strains.

That this third hypothesis could encompass all of our data is not too surprising since it was deliberately tailored to fit data already available. However, this hypothesis did not have to remain a post-hoc rationalization. It could be put to independent test, since it generated new predictions about strain differences in brain biochemistry about which no observations existed. Specifically, this hypothesis demanded that the S_1 strain have greater ACh than the S_3 strain, while smaller differences in ACh (or none at all) should exist within the RDH–RDL pair and the RCH–RCL pair. We ventured this prediction soon after we had arrived at our third hypothesis but before we had the means for its testing (Krech, et al., 1958; Rosenzweig, et al., 1958c). The collaboration of James Crossland, of St. Andrews University, as USPHS consultant in our laboratories during the summer of 1959, allowed us to carry out the ACh assays required for the testing of these predictions.

To preserve the highly labile ACh, it was necessary to work with frozen tissue, and this made it extremely difficult to obtain separate cortical and subcortical samples. The ACh assays were therefore carried out on tissue samples which included all of the brain anterior to the pons and cerebellum (the pons, cerebellum, and medulla being excluded). The analytical techniques are summarized in Bennett, Crossland, Krech, and Rosenzweig (1960). ChE was then analyzed on brain samples comparable to those employed for the ACh assays.

The results of these analyses are presented in Table 2. Two major

points can be seen from the ACh data in the third column of the table: (*a*) The S_1 strain has a significantly greater concentration of ACh than does the S_3 strain, as we had predicted. The difference is about 14% and is significant at the .001 level of confidence. These data appear to be the first evidence for strain differences in ACh concentrations. (*b*) Neither of the pairs of Roderick strains shows a difference this large, again following the prediction. The RCH rats show a concentration of ACh about 9% greater than that of the RCL rats, and this difference is significant at the .05 level of confidence. The difference between the RDH and RDL strains is less than 1% and does not approach significance.

The data for ChE activity (presented here for the first time) are comparable to those which we have previously reported for somewhat different tissue samples. There is a relatively large and highly significant difference within each pair of strains. The ChE activity of the S_1 strain exceeds that of the S_3 strain by 10%, the RCH value exceeds the RCL value by 23%, and the RDH value exceeds that of the RDL by 19%.

These results appear to confirm our new assumption that ACh and ChE are independent genetically. The greatest difference in ACh concentrations occurs between the S_1 and S_3 strains, while these strains show the least difference in ChE activity. Conversely, the RDH and RDL strains show no difference in ACh concentrations, although they differ markedly in ChE activity. In other words, the program of selection for ChE activity was specific in its effects and had little influence on ACh concentrations. It is interesting to note that we have here an instance of an enzyme and its substrate being under independent genetic control. In any event, it is clear that, among strains, ChE activity does not afford a valid index of the concentration of ACh for the brain, as sampled in our assay.

Since the ACh and ChE measures may be largely independent among strains, it is necessary to determine both in order to characterize the functioning of the biochemical system of which they are components. (Information about the third member of the system, ChA, may be of further value, and we are now attempting to obtain it.) The S_1 strain exceeds the S_3 strain significantly in both ACh concentration and ChE activity. As we have pointed out earlier, this combination should make for more rapid and more certain synaptic transmission in the S_1 strain. This more efficient transmission should, in turn, make for superior learning capacity. In the case of the paired Roderick strains, the greater ChE activity of the high-ChE lines is not accompanied by an equally great relative increase in ACh concentration. This, we have suggested, should bring about more rapid destruction and less persistent action of released ACh at the synapse in the Roderick high-ChE strains, and consequently should result in less certain synaptic conduction.

It will be seen that we have restricted our comparisons to strains that are derived from a common foundation stock (S_1 vs. S_3, RCH vs. RCL, RDH vs. RDL). This is done because we consider ACh functioning to be only one of many factors which determine learning capacity. Within paired strains, the other factors determining learning capacity are likely to be in common, only the selected physiological characteristic and factors closely related to it differing

The comparison between paired strains, while still somewhat risky, reduces the ambiguity of conclusions. If strains deriving from different foundation stocks were compared, the attempt to relate behavioral differences to any one of the large variety of physiological differences existing between the strains would be perilous.

The last column of Table 2 presents ratios of the ACh and ChE values for each strain. While some of the differences within pairs of strains are quite small, it is interesting to note that within each pair, the strain with the greater ratio is the one which we have previously seen (Table 1) to be superior in learning behavior. These differences in time required for the transmitter substance to be hydrolyzed by its enzyme thus accord with the behavioral data. That is, for the strain that learns better, the ACh continues to function for a somewhat longer time at the synapse before being hydrolyzed. Nevertheless, we do not wish to put much emphasis upon the ratio, for two reasons: (*a*) The ratio does not preserve any indication of the absolute levels of ACh and ChE, and these levels are probably of considerable importance. If two strains (or individuals) had identical ratios, but one was high in both ACh and ChE while the other was low in both, we would expect to find the first superior to the second in problem-solving ability. This appears to be essentially the case with the S_1 and S_3 strains, whose ratios are quite similar. (*b*) Use of a ratio might imply that all of the ACh could be acted upon by all of the ChE. Actually we know that both compounds show distinct patterns of distribution among discrete brain regions, and there is also reason to believe that different regions are of decidedly different importance for learning. Thus these ratios can provide only gross estimates of the relative length of time required for destruction of the transmitter substance at synapses involved in learning.

Current and Projected Studies

The third hypothesis, like its predecessors, is only provisional and is at present being subjected to further tests. In this section we will indicate further chemical and physiological tests and genetic selection studies, in progress and projected.

The chemical tests are concerned with the validity of our conclusion that ACh concentration and ChE activity are largely independently controlled. The large brain sample which we used in assaying for ACh may have obscured positive relations between ACh and ChE activity between strains for *defined subregions of the brain that are important in learning*. If such positive relations exist, they would provide evidence against our hypothesis. This makes it desirable to obtain values for ACh and ChE in subregions of the brain. We are now doing ChE analyses on such samples. Since, as we have indicated, we cannot take precisely defined sections when the brain has been frozen for ACh assay, we are attempting to determine, in collaboration with Crossland, whether ChA activity can be used as an index to ACh concentration. Since analysis of ChA activity does not require a frozen brain, the assay for ChA can be done on precise brain samples.

A physiological test frequently used to provide a measure of cortical excitability consists of delivering electrical shocks through the head and determining the threshold for seizures. With this minimal thresh-

old seizure test, the S_1 strain has been found to have a lower threshold and therefore greater cortical excitability than the S_3 strain (Woolley, Rosenzweig, Krech, Bennett, & Timiras, 1960). This is, of course, consistent with our chemical findings about the differences between these two strains, and we plan to subject the Roderick strains to this electrophysiological test. We should expect the RCH to have a somewhat higher threshold than the RCL strain, and the RDH somewhat higher than the RDL.

A genetic experiment is in progress to test our third hypothesis. This experiment is essentially patterned after Tryon's experiment but with concurrent analyses of brain biochemistry. It is being conducted by Richard Olson, a graduate student in the genetics department. From each of two heterogeneous foundation stocks, one line is being developed for superior learning and one for inferior learning in the Lashley III Maze. In each generation, the brains of half of the animals are being assayed for ACh concentration and the other half are being analyzed for ChE activity. The chemical measurements will not enter into the selection program which is based entirely upon the behavioral scores. The aim of the experiment is to determine whether progressive selection for learning capacity will entail concurrent changes in ACh and ChE similar to the differences found between the descendants of Tryon's maze-bright and maze-dull strains.

Another genetic experiment which we hope to undertake soon involves selection of rats for ACh concentration. Such selection should bring about a compensatory change in ChE activity. (We have summarized elsewhere evidence that increases and decreases of ACh metabolism lead to similar changes in ChE activity—Krech, et al., 1960). This genetic selection for high ACh concentrations alone should yield a strain that is high in *both* ACh and ChE, the change in ChE being an instance of induced (nongenetic) enzymic response. Similar genetic selection for low ACh concentration should yield a strain low in *both* ACh and ChE. We would predict that the former strain will be superior to the latter strain on learning tests.

Conclusion

In the attempt to test our general hypothesis—that variation in brain chemistry is a major determinant of variation in adaptive behavior among normal individuals—we have rejected two specific hypotheses and are now testing a third one. Testing each hypothesis has produced new observations about brain chemistry, about behavior and about relations between brain chemistry and behavior.

Some of the main findings of our laboratories are these: (*a*) Strains and individuals may differ significantly in brain ACh concentration and ChE activity. (*b*) Rats can be bred selectively for brain ChE activity. (*c*) Selective breeding for an enzyme (ChE) may have little or no effect on the concentration of its substrate (ACh). (*d*) Measures of brain chemistry (ACh concentration, ChE activity, LDH activity, percent protein) show relatively little variation among individuals, compared to the variability found in other organs. (*e*) The pattern of distribution of ChE activity among brain regions differs from that of LDH activity, or percent protein. (*f*) ChE activity is a general characteristic of the brain since values for one region show high

correlations with those for other regions. (*g*) Pentobarbital sodium causes rats to fixate on one attempted solution of a problem rather than to try other possible solutions. (*h*) Tryon's genetic selection program has been shown to have had rather general results, since the S_1 strain is superior to the S_3 strain on three behavioral tests, none of which was employed in the original selection. (*i*) Strains bred for high cortical ChE activity tend to be inferior on behavioral tests to strains from the same foundation stocks bred for low ChE activity. (*j*) In strains descended from crosses between the S_1 and S_3 strains, error scores and cortical ChE tend to be correlated positively.

The relations between variation in brain chemistry and variation in problem-solving ability for three paired strains (S_1 vs. S_3, RCH vs. RCL, RDH vs. RDL) and for a crossed strain (K) can all be encompassed by our current hypothesis. As tests of this hypothesis lead to discovery of more facts about brain chemistry, behavior, and their relations, undoubtedly further progress in theory will be needed to summarize and to stimulate further progress in experimentation.

REFERENCES

Bennett, E. L., Crossland, J., Krech, D., & Rosenzweig, M. R. Strain differences in acetylcholine concentration in rat brain. *Nature*, 1960, **187**, 787–788.

Bennett, E. L., Krech, D., Rosenzweig, M. R., Karlsson, H., Dye, N., & Ohlander, A. Cholinesterase and lactic dehydrogenase activity in the rat brain. *J. Neurochem.*, 1958, **3**, 153–160. (a)

Bennett, E. L., Rosenzweig, M. R., Krech, D., Karlsson, H., Dye, N., & Ohlander, A. Individual, strain and age differences in cholinesterase activity of the rat brain. *J. Neurochem.*, 1958, **3**, 144–152. (b)

Bennett, E. L., Rosenzweig, M. R., Krech, D., Ohlander, A., & Morimoto, H. Cholinesterase activity and protein content of rat brain. *J. Neurochem.*, in press.

Burkhalter, A., Jones, M., & Featherstone, R. M. Acetylcholine-cholinesterase relationships in embryonic chick lung cultivated *in vitro*. *Proc. Soc. Exp. Biol. Med.*, 1957, **96**, 747–750.

Crossland, J. Chemical transmission in the central nervous system. *J. Pharm. Pharmacol.*, 1960, **12**, 1–36.

Krech, D., Rosenzweig, M. R., & Bennett, E. L. Dimensions of discrimination and level of cholinesterase activity in the cerebral cortex of the rat. *J. comp. physiol. Psychol.*, 1956, **49**, 261–268.

Krech, D., Rosenzweig, M. R., & Bennett, E. L. Brain chemistry and adaptive behavior in the rat. *Science*, 1958, **128**, 1146.

Krech, D., Rosenzweig, M. R., & Bennett, E. L. Inter-hemispheric effects of cortical lesions on brain biochemistry. *Science*, 1960, **132**, 352–353.

Krech, D., Rosenzweig, M. R., & Bennett, E. L. Effects of environmental complexity and training on brain chemistry. *J. comp. physiol. Psychol.*, in press.

Krech, D., Rosenzweig, M. R., Bennett, E. L., & Krueckel, B. Enzyme concentrations in the brain and adjustive behavior-patterns. *Science*, 1954, **120**, 994–996.

Krechevsky, I. The genesis of "hypotheses" in rats. *U. Calif. Publ. Psychol.*, 1932, **6**, 45–64.

Krechesvky, I. Brain mechanisms and "hypotheses." *J. comp. Psychol.*, 1935, **19**, 425–448.

McGaugh, J. L. Some neurochemical factors in learning. Unpublished doctoral thesis, University of California, Berkeley, 1959.

McGaugh, J. L., & Petrinovich, L. The effect of strychnine sulphate on maze-learning. *Amer. J. Psychol.*, 1959, **72**, 99–102.

McLennan, H., & Elliott, K. A. C. Effects of convulsant and narcotic drugs on acetylcholine synthesis. *J. Pharmacol.*, 1951, **103**, 35–43.

Moroz, M. Effect of pentobarbital sodium on the behavior of rats in the Krech hypothesis apparatus. *J. comp. physiol. Psychol.*, 1959, **52**, 172–174.

Nachmansohn, D. Sur l'action de la strychnine. *CR Soc. Biol., Paris*, 1938, **129**, 941–943.

Neilands, J. B., & Cannon, M. S. Au-

tomatic recording pH instrumentation. *Analyst. Chem.*, 1955, **27**, 642–651.

Peirce, J. T. Cerebral cholinesterase activity and "hypothesis" behavior in a problemless apparatus. *J. comp. physiol. Psychol.*, 1959, **52**, 168–171.

Roderick, T. H. Selection for cholinesterase activity in the cerebral cortex of the rat. *Genetics*, in press.

Rosenzweig, M. R., Krech, D., & Bennett, E. L. Effects of pentobarbital sodium on adaptive behavior patterns in the rat. *Science*, 1956, **123**, 371–372.

Rosenzweig, M. R., Krech, D., & Bennett, E. L. Brain chemistry and adaptive behavior. In H. F. Harlow and C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958. Pp. 367–400. (a)

Rosenzweig, M. R., Krech, D., & Bennett, E. L. Brain enzymes and adaptive behaviour. In, *Ciba Foundation Symposium on Neurological Basis of Behaviour*. London: Churchill, 1958. Pp. 337–355. (b)

Rosenzweig, M. R., Krech, D., & Bennett, E. L. Heredity, environment, brain biochemistry, and learning. In, *Pittsburgh Symposium on Current Trends in Psychology*, March 12, 1959.

Rosenzweig, M. R., Krech, D., Bennett, E. L., & Longueil, C. L. Strain differences of rats in behavior and brain chemistry. *Amer. Psychologist*, 1958, **13**, 405. (Abstract) (c)

Russell, R. W. Effects of reduced brain cholinesterase on behavior. *Bull. Brit. Psychol. Soc.*, 1954, **23**, 6. (Abstract)

Tower, D. B. The neurochemical substrates of cerebral function and activity. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958. Pp. 285–366.

Tryon, R. C. Genetic differences in maze learning ability in rats. *Yearbk. Nat. Soc. Stud. Educ.*, 1940, **39**, Part I, 111–119.

Woolley, D. E., Rosenzweig, M. R., Krech, D., Bennett, E. L., & Timiras, P. S. Strain and sex differences in threshold and pattern of electroshock convulsions in rats. *Physiologist*, 1960, **3**, 182.

(Early publication received June 20, 1960)

Psychological Bulletin
1960, Vol. 57, No. 6, 493–513

PSYCHOLOGICAL AND RELATED CHARACTERISTICS OF SMOKERS AND NONSMOKERS

JOSEPH D. MATARAZZO AND GEORGE SASLOW
University of Oregon Medical School

Considering all the evidence available now, the writers themselves cannot doubt that an extremely high degree of association exists between cigarette smoking and lung cancer. This association has been reported independently by many different investigators who cannot all possibly have committed the very same errors. The evidence appears to support strongly the hypothesis that cigarette smoking is a major causative factor in lung cancer. The present status of the cause-effect controversy is described most cogently in the articles by Backett (1958), Hammond (1958), Little (1957), and Rutstein (1957), the review by the Study Group on Smoking and Health (1957), the book by Northrup (1957), and the public statement by Leroy E. Burney.[1] As Rutstein (1957) has pointed out, the general health problem posed by excessive cigarette smoking is of sufficient magnitude to warrant consideration *now* of what preventive measures may be socially possible and desirable. Although data now clearly indicate that any reduction in the number of cigarettes smoked by an individual reduces the health hazard (Hammond, 1958, Figure 6, p. 338), the percentage of smokers who have given up smoking is quite low, falling somewhere between 10 and 18.9% (Haenszel, Shimkin, & Miller, 1956, p. 24; Hammond & Percy, 1958, p. 2956).

In view of the current widespread interest in the smoker, it has seemed to us wise to review the literature on what is known about the psychological, personal, social, and situational characteristics of smokers and nonsmokers. Few systematic studies exist which are concerned directly with this issue. Some of the pertinent findings are summarized below. Two of the best studies apparently have not yet received wide attention.

The first of these was a study by Haenszel et al. (1956). As a supplement to the United States Bureau of the Census Current Population Survey (CPS) for February 1955, smoking histories were collected from approximately 40,000 men and women 18 years of age and over. Survey data were adjusted to include unsurveyed population groups (Armed Forces, teenagers under 18, and institutional population).

It was possible with apparently a fair degree of accuracy for the authors to generalize from this representative sample of 40,000 subjects to the smoking habits of the total adult population of the United States (some 105 million adults). Adjusted estimates largely based on data from this 1955 survey, the 1952 survey of some 187,000 older men made by Hammond and Horn for the American Cancer Society (1954, 1958), and a variety of other statistics-gathering agencies (Haenszel, et al., 1956) indicate that the total number of smokers in the United States and in overseas forces as of early 1955 was close to 60 million. Of these 54 million smoked regularly (everyday) and the rest occasionally (not everyday). The number of regular and occasional smokers in

[1] Surgeon General, Public Health Service, United States Department of Health, Education, and Welfare, July 12, 1957.

TABLE 1

AGE AND SMOKING[a]

(Regular Cigarette Smokers in the United States by Age and Sex)

Age	Males			Females			Both Sexes		
	Total *N*	Smokers	%	Total *N*	Smokers	%	Total *N*	Smokers	%
18–24	5,405,000	2,487,000	46	7,460,000	2,346,000	31	12,865,000	4,833,000	38
25–44	21,820,000	12,378,000	57	23,680,000	8,933,000	38	45,500,000	21,311,000	47
45–64	16,034,000	7,250,000	45	16,695,000	3,295,000	20	32,729,000	10,545,000	32
65 and over	6,322,000	1,296,000	20	7,261,000	379,000	5	13,583,000	1,675,000	12
Total	49,581,000	23,411,000	47	55,096,000	14,953,000	27	104,677,000	38,364,000	37

[a] Abstracted from Table 2, p. 57, Haenszel, Shimkin, and Miller (1956).

the civilian, noninstitutional population, 18 years and over, was about 51½ million—nearly half of this population segment. The number of male smokers totaled more than 31 million who smoked regularly and 4 million who smoked occasionally. The number of female smokers totaled 14 million regular and 2 million occasional (pp. 11–15).

The numbers shown in the tables presented below vary slightly from these, due to the assumptions made, especially regarding the unsurveyed (institutional, etc.) groups (Sackrin & Conover, 1957, p. 1).

The second study, by Sackrin and Conover (1957), was a direct outgrowth of the study by Haenszel et al. and was based on a subsample of 18,000 of the 40,000 cases for whom Haenszel et al. had smoking histories and on whom it was possible for Sackrin and Conover to obtain additional information on income. The pertinent results from both these studies are given in terms of Variables Age through Income below.

A third study by Lilienfeld (1959) also appears to be a major contribution in this area. A representative sample of 4,456 adults (18 years of age and over) in Buffalo, New York, was studied by face-to-face interview and multiple-choice questionnaire to determine whether cigarette smokers differed from nonsmokers with respect to emotional status and other selected characteristics. From this sample, 903 cigarette smokers were matched with 903 nonsmokers with respect to age, sex, race, and social status. The chief findings of these three major studies, as well as of numerous others, will now be summarized under their appropriate headings.

PERSONAL-SITUATIONAL VARIABLES

Age and cigarette smoking. The data in our Table 1 have been abstracted from Haenszel et al. (1956). Only a very few males below the age of 10 and probably no females in this very young age group smoke. For Americans, smoking apparently begins in the early and late teens. Thus no data are given for the ages 0–17 years. While smoking patterns have changed markedly in the past 50 years, and are continuing to change, in the year 1955 age (beyond adolescence) had a curvilinear relationship to smoking. Table 1 shows that for the two sexes combined, in the younger age ranges (18–24), approximately one out of three (38%) individuals were smokers; this increases to one out of two (47%) individuals in the early middle age group (25–44); and then declines again for the older group (65 and over) to about one out of every eight (12%). Additional data on age and smoking are to be found in other sources (Hammond & Horn, 1954, 1958; Heath, 1958; McArthur, Waldron, & Dickinson, 1958; Sackrin & Conover, 1957).

TABLE 2

SEX AND SMOKING[a]

(Smokers in the United States Aged 18 and Over)

	Males		Females	
	N	% of Total	*N*	% of Total
Total Group[b]	49,581,000		55,096,000	
Regular Smoker (cig., cigars, and/or pipes)	33,566,000	68	14,953,000	27
Smoked Occasionally	2,448,000	5	2,259,000	4
Never Smoked	10,704,000	22	35,785,000	65
No Data	2,863,000	6	2,099,000	4
Regular Cigarette (only) Smokers	23,411,000	47	14,953,000	27

[a] Abstracted from Table 2, p. 57, Haenszel, Shimkin, and Miller (1956).
[b] This total does not equal the sum of the number shown below because some smokers, primarily among males, often smoke pipes and cigars, as well as cigarettes, and thus any given individual may be represented twice, once as a Regular Smoker and once as a Regular Cigarette (only) Smoker.

Sex and smoking. As is suggested in Table 1 and summarized in Table 2, fewer women (27%) over age 18 than men (68%) over 18 smoke regularly. In women, smoking is almost exclusively confined to cigarettes, whereas with men 68% smoke regularly in one form or another, while only 47% regularly smoke cigarettes exclusively. Although not shown in our Table 2, the data gathered by Haenszel et al. further showed that in the year 1955, in the 18–24 age group, 34% of males had started smoking by age 18, and 50% had begun by age 24. For females in the 18–24 age group, 16% had started smoking by age 18, and 36% had begun by age 24 (Haenszel, et al., 1956, p. 56).

Although the figures on percent smokers for young males have changed little in the past 50 years, young women have shown a substantial increase. Whereas few women now in their 40's and 50's smoked at an early age, 20% of the women now in their 20's were regular smokers by the age of 18.5 (Haenszel, et al., 1956, p. 17). Thus as is the case for sexual habits (Kinsey) smoking habits of women born in the past several decades are very different from those of their mothers and grandmothers. A progressive loosening of "moralistic" attitudes in the past 50 years has been suggested as a contributing factor.

Evidence that the relationships shown in Tables 1 and 2 between smoking and age and sex are not unique to the inhabitants of the United States comes from a recently completed study of the smoking habits of the entire population of Israel (Kallner, in press). Except for minor differences, the age and sex relationships were similar.

Race and cigarette smoking. The United States data of Haenszel et al. (1956, pp. 36–38) permit the following conclusions: The age and sex differences noted above hold for nonwhites as well as whites. In the United States, there is no difference between the percentage of whites and the percentage of nonwhites who smoke. However, quantity of smoking differs between whites and nonwhites. Among smokers, there is a

significantly greater percentage of heavy cigarette smokers among whites (more than one pack a day) than there is among nonwhites. The percentage of white males is 13.3 and for nonwhite males 6.9. A similar difference is found for females. Indirect evidence suggests an economic factor is involved in these differences.

Although Lilienfeld's (1959) smokers and nonsmokers were matched for race, his data show that, despite the fact that their own birthplace did not differentiate smokers from nonsmokers, the parents of nonsmokers were found to be more often foreign-born than those of smokers ($p<.001$). The author states that the implication of this difference is not clear.

Marital status and cigarette smoking. The data from Haenszel et al. (1956, pp. 44–46) permit these conclusions: For all ages, and for both sexes, there is a greater percentage of smoking among divorced and widowed individuals than there is among both married and never-married individuals. Among single individuals (never-married) under age 45, one finds a smaller percentage of smokers than among married individuals of comparable age. There are more single individuals who have never smoked than there are married individuals who have never smoked (in the 18–45 age group). There is a slight reversal of this trend in the age group over 55.

Lilienfeld's (1959) results bear out one aspect of the findings of Haenszel et al., in that he found that, although present marital status does not differentiate smokers from nonsmokers, the smokers had previously married significantly more often ($p<.001$).

Occupation and cigarette smoking. Among males a slightly larger proportion of the unemployed than the employed smoke. Professional and technical workers, although having the highest incomes, have a low smoking proportion (also see section below on Income). There seems to be a complex relationship between social class and smoking patterns. White collar groups (professional workers, managers, etc.) have fewer smokers than are found among craftsmen, foremen, salespersons, operatives, and similar groups (for further data on this "social class" variable see section below on Socioeconomic status). It was found that military life is associated with a higher frequency of smoking for all age groups (25–65 and over); veteran age groups were found to have more smokers than the nonveteran age groups (Haenszel, et al., 1956, pp. 38–44; Sackrin & Conover, 1957, p. 5).

In the sample of 1,806 Buffalo adults, Lilienfeld found that smokers change jobs significantly more often than do nonsmokers ($p<.001$).

Additional (longitudinal research) findings on the relationship between occupation and cigarette smoking have been reported recently by Heath (1958) and McArthur et al. (1958). These data are shown below in Table 5.

Urban-rural residence and cigarette smoking. Present urban-rural residence differentiates smokers from nonsmokers sharply. There is a smaller percentage of smokers of both sexes and at all ages in the rural farm population than in either the rural nonfarm or in the city population. Rural nonfarm persons resemble closely urban dwellers in their smoking habits. Among males there is little or no variation in smoking patterns from one region of the United States to another. On the other hand, the 1955 Census results show that cigarette smoking is more prevalent among women in northeastern and

TABLE 3

INCOME AND REGULAR CIGARETTE SMOKING[a]

Males		Females	
1954 Income	% Smokers	% Smokers	1954 Income
Less than $1,000	39	19	Less than $1,000
$2,000–$7,000	56–60	28–33	$2,000–$4,000
$7,000 and Over	50	32	$4,000 and Over
		23	No Money Income (Homemakers)

[a] Abstracted from Sackrin and Conover (1957, pp. 2–3).

western states than elsewhere in the United States (Haenszel, et al., 1956; Sackrin & Conover, 1957).

Income and cigarette smoking. The main findings of Sackrin and Conover (1957), from their study of 18,000 Americans, have been abstracted in Table 3. It is clear from the results in the table that *regular* cigarette smoking is, in fact, related to reported annual income.

> The proportion of males who smoke regularly increases from 39% of the men whose annual income is less than $1,000, to 56–60% of those in four income brackets from $2,000 to $7,000. For men receiving $7,000 and over, the proportion of regular cigarette smokers drops to a little over 50% (p. 2).
>
> Less than a fifth of the women whose personal incomes are less than $1,000 a year smoke regularly, but the proportion increases to about a third for women receiving incomes of $3,000 or more. About a fourth of the women who receive no personal income (largely home-makers) smoke cigarettes (p. 5).
>
> Income appears to have a greater effect on rates of smoking than on the percentage of regular smokers, although differences associated with age and other population characteristics were noted also. The majority of all regular cigarette smokers smoke from 10 to 20 cigarettes daily, generally regardless of age or income level. Enough men exceed this range to bring the overall daily average to 21–22 cigarettes for male smokers. The daily average for women . . . is 14–16 cigarettes daily (p. 8).

Socioeconomic status and cigarette smoking. From our own department, Allen (1958) reported the smoking behavior of three different groups. The first was a group of 40 psychiatric patients from the Massachusetts General Hospital whom we studied by means of individual psychiatric interview of a special standardized kind and by a battery of individually administered objective and projective personality tests. Of this group, 31 were cigarette smokers, 9 were nonsmokers. The second group consisted of 114 female student nurses from a school of nursing located in the northwest, of which 50 were cigarette smokers and 64 were nonsmokers. The third group included 140 male and female university undergraduate students, also from the northwest. Seventeen of the females were smokers, 31 were nonsmokers; 54 of the males were smokers, 38 were nonsmokers. Characteristics of these groups are given in Allen (1958) and Matarazzo, Matarazzo, Saslow, and Phillips (1958).

The measure of socioeconomic status used in each of the three groups was the Hollingshead scale (Hollingshead & Redlich, 1958), an index which combines own (or parents') education and present occupation to yield a single socioeconomic status score.

As can be seen in the first row of Table 4 below, no relationship was found between socioeconomic status

and smoking in any of the three groups. Thus, studying a total of 294 individuals of widely varying ages and a considerable range in socioeconomic status yielded the finding that smoking seems to be unrelated to this measure of socioeconomic status. However, a study by McArthur et al. (1958), described below, indicates that, for Harvard undergraduates, "nonsmokers tend to be lower-middle class in origin, upwardly mobile, earnest young men . . . " (p. 274). These authors (p. 269) cite a survey published in England in 1948 which "suggested that nonsmoking was commoner among men of the English middle class while heavy smoking was commoner among English working-class men." These results appear to conflict with our own findings. Furthermore, our own findings with the Hollingshead measure of socioeconomic level seem to conflict with the finding by Sackrin and Conover of a relationship between earned annual income and smoking. However, when it is remembered that education and occupation (the Hollingshead measure) do not correlate highly with annual income, and that in all probability, neither correlates highly with the McArthur et al. measure, it is apparent that better indices of socioeconomic status are required or, in their absence, a study employing the available indices with the same subjects might shed some light on these inconsistencies.

Education and cigarette smoking. The most direct results bearing on this relationship were given by Lilienfeld (1959, p. 277). His study showed that adult smokers and nonsmokers do not differ significantly in final number of years of schooling completed. There were as many smokers as nonsmokers who had had no schooling, or who had attended college, etc.

For people still in school, since age and number of years of school completed are obviously correlated up to about age 25, it can be concluded that up through high school, college, and professional school there is an increase in the percentage of smokers as grade level of students still in school increases. A study of 6,374 college students in 11 Texas colleges (aged 15–39) by Kirchoff and Rigdon (1954, p. 296) showed just such an increase of 30% to 63% during the four college years. A study by Horn, Courts, Taylor, and Solomon (1959), on the smoking habits of the 22,000 high school students in Portland, Oregon, extends these observations backward into the high school years. For boys, smoking increased from 14.5% of the total in the freshman year to 35.4% in the senior year. The comparable increase for girls was from 4.6% in the freshman year to 26.2% in the senior year. Raven (1957) provides some anecdotal observations on the smoking habits of English schoolboys.

A study made of a group of college students at Antioch by J. R. Earp (1936) showed that of 177 students who smoked, 57% failed to graduate; of 176 nonsmokers 31.8% failed to graduate. Vallance (1940–45, p. 139) correctly points out that this finding is merely a correlation and reveals nothing about smoking being the cause of college failure; individuals who earn poor grades might be the ones who take up smoking.

A study of smokers and nonsmokers as related to achievement and various personal characteristics made by R. M. Lynn (1948) showed that adolescent boys who do not

TABLE 4
MEANS AND RANGES OF SMOKERS AND NONSMOKERS

Variable	Psychiatric Patients ($N=40$)		Student Nurses ($N=114$)		University Undergraduates ($N=140$)			
			Females		Females		Males	
	Non-smokers ($N=9$)	Smokers ($N=31$)	Non-smokers	Smokers	Non-smokers	Smokers	Non-smokers	Smokers
Socioeconomic Index:								
Mean	60.2	57.9	48.6	46.2	43.6	42.1	49.8	45.4
Range	44–73	14–77	11–73	11–77	11–71	11–73	11–77	11–73
IQ:								
Mean	93.6	98.8	117.6	118.4	110.9	109.2	107.9	109.2
Range	77–109	79–129	103–130	103–129	87–129	92–122	84–130	89–131
Anxiety level:								
Mean	28.9	25.9	12.3	14.8*	12.0	15.3	11.0	14.7*
Range	13–39	6–45	3–26	3–34	5–28	6–45	2–30	1–33
Psychosomatic Symptoms:								
Mean	12.1	13.9	6.3	8.2*	3.7	6.1	3.3	3.9
Range	2–23	1–44	0–22	0–18	0–14	0–18	0–12	0–19
Cups of Coffee:								
Mean	2.8	4.2	0.9	2.6**	1.5	2.7	1.0	3.5**
Range	0–8	0–15	0–6	0–10	0–10	0–6	0–6	0–12
Liquor Score:								
Mean	1.22	2.06	1.0	1.2*	1.3	1.5	1.5	2.2**
Range	1–2	1–6	1–2	1–2	1–2	1–2	1–5	1–6

* Mean differences significant at the .05 level.
** Mean differences significant at the .001 level.

smoke, on the average will gain more weight, make higher grades in school, fail less often, cause less disciplinary trouble, make better scores on psychological tests, be troubled less with respiratory diseases, than the occasional smoker and the habitual smoker. The study showed also that smoking and poor scholarship do not always go together, in that scholastic averages according to age favor the nonsmokers in some cases and the smokers in others.

Participation in sports and cigarette smoking. Lilienfeld (1959) found that smokers had participated to a greater extent than nonsmokers in certain major sports ($p<.01$), had participated more frequently in a miscellaneous grouping of other nonspecified sports ($p<.05$), and, at the time of the interview, were participating in a greater number of sports ($p<.02$). What relationship, if any, there is to Heath's finding that smokers had more combat duty in World War II than nonsmokers, and Ianni's finding of more driving accidents among smokers, is unclear. These studies will be described below.

Driving accidents and cigarette smoking. Data for this variable were

made available to us in a personal communication by Ianni and Boek[3] (from Russell Sage College) who found, in a well-controlled study, that in a group of 161 drivers who had just been in a driving accident, there was a higher frequency (76%) of current cigarette smokers than in a control group of 196 nonaccident drivers selected from the same driving population which had only 54% current cigarette smokers (p computed from their data is $<.001$). The number of ex-smokers was 13% in both the accident and nonaccident control group.

Psychological Variables

IQ and cigarette smoking. The three groups studied by us provided data for this variable. IQ was determined by the Wechsler Adult Intelligence Scale for the 40 patients, and by the Otis Test of Mental Ability for the two younger age groups shown. As can be seen in Table 4, and despite the adequacy of the IQ range covered by our three samples (77 to 131), smokers do not differ from nonsmokers on this variable. Thus it can be concluded that IQ is probably not related to smoking. The relationship between this finding and the previously cited finding that nonsmokers tend to earn better grades is unclear and points to the need for further research. However, it should be remembered that, as pointed out by Super (1949, p. 90), "the correlation between intelligence tests and *grades* is not especially high . . . the modal r's being .30 to .50."

Anxiety and cigarette smoking. The three groups investigated by us were studied on this variable also. The measure of anxiety used in each case was the Taylor Manifest Anxiety Scale, a reliable and reasonably well validated 50-item questionnaire designed to reflect conscious anxiety (Matarazzo, Guze, & Matarazzo, 1955; Taylor, 1953; Taylor & Spence, 1952).

As is seen in Table 4, for the 40 psychiatric patients the mean anxiety score for both the nonsmokers (28.9) and smokers (25.9) is very high ($p<.001$) relative to the mean anxiety scores in the two samples of normals (means ranging from 11.0 to 15.3), both smokers and nonsmokers. The higher anxiety scores for the 40 psychiatric patients are very similar to the scores reported by us for two other psychiatric populations studied at the Washington University School of Medicine Outpatient Clinic (Matarazzo, et al., 1955); while the lower mean scores for the two normal groups also are exactly in the range typically found for unselected groups of normals (Matarazzo, et al., 1955; Taylor, 1953).

The results shown in Table 4 make it clear that, for young normal subjects, smokers have higher anxiety scores than nonsmokers. While the differences between means are not great, they nevertheless reach statistical significance ($p<.05$) in two of the 3 samples, and show a similar trend in the third (female university undergraduates). Relative to the nonsmokers in these two normal groups, the higher anxiety scores among the (normal) smokers, place the smoker at a point along the anxiety continuum which is closer to that characteristic of psychiatric patients.

That a similar finding of a higher anxiety score among smokers, rela-

[3] Ianni, F. A., & Boek, W. A study of the relationship between motor vehicle accidents and certain characteristics of drivers. Unpublished manuscript, 1958.

tive to nonsmokers, was not obtained in the psychiatric population may well be a reflection of the higher overall level of anxiety found in both smokers and nonsmokers in this population. The small number (nine) of nonsmokers in our psychiatric population does not permit us to establish this point.

Psychological tension and cigarette smoking. In an interesting study of 63 (normal) medical-surgical male patients and 32 psychiatric male patients in a Rhode Island Veterans Administration Hospital, Lawton and Phillips (1956) used the Cornell Medical Index and another specially devised questionnaire. The CMI is a well-validated 195-item questionnaire (Brodman, Erdmann, & Wolff, 1949) covering a wide range of somatic and psychological symptoms, not unlike the Taylor MAS. The authors concluded that among their normals, "Heavy Smokers," when compared to "Moderate Smokers," showed a greater number of signs of psychological tension (12 p values from .01 to .05). They concluded that

> the present data appear to indicate a very real tendency for this present group of heavy smokers to exceed the group of moderate smokers and abstainers in various indices relating to presence of "nervous" traits. In both number of somatic and psychological complaints, the heavy smokers seem to resemble the emotionally disturbed individual more than do the moderate smokers. In the manner in which they are willing to describe themselves [on the Adjective Check List of the second questionnaire], the heavy smokers . . . [describe themselves as] less agreeable, happy, and relaxed . . . [and specifically rate themselves more often as] nervous and grouchy (p. 401).

In addition, it was found that there was a significantly higher percentage of Heavy Smokers as against Moderate Smokers in the sample of 32 psychiatric patients than in these 63 nonpsychiatric medical-surgical patients.

Thus two conclusions from Lawton and Phillips' study are indicated. First, that Heavy Smokers among the normals are more like the psychiatric patients by these various indices, and second, that psychiatric status is associated with a higher frequency of smoking.

The first conclusion bears out our own observations on anxiety, discussed above. The lack of comparability on the age factor between our own two normal samples and our psychiatric sample precludes any useful test of the second conclusion drawn from Lawton and Phillips' data.

Suggestibility and smoking. Stimulated by the statement made by 29 out of 35 college student smokers that they had been influenced to take up smoking by their friends, Vallance (1940–45) studied the suggestibility of smokers versus nonsmokers. Using Hull's body sway technique, he found among Miami University students (25 smokers and 22 nonsmokers) just the opposite; i.e., that the smokers were less suggestible than were the nonsmokers on this measure. With the subject having his eyes closed while standing erect, the suggestion was made by the experimenter that the subject was to imagine that he was falling forward. Under this set of suggestions the smokers swayed only 3.83 cm. while the nonsmokers swayed 5.27 cm. Other measures of suggestibility conceivably could show different results, since it is well-known that Hull's measure is not highly correlated with other measures of suggestibility.

There are such strong a priori reasons to think that parental and peer influences are strong factors in the initiation of the smoking habit

that one certainly should not conclude from Vallance's negative results with a crude laboratory measure of suggestibility that the issue is closed. As a matter of fact, the study by Horn et al. (1959) of 22,000 high school students in the Portland, Oregon, metropolitan area shows just such influences. In addition to many other important findings, their data reveal that: (*a*) the percentage of smokers is highest among children of families in which both parents smoke cigarettes, and lowest in families in which neither parent has been a smoker; (*b*) the percentage of smokers is higher in Catholic parochial schools than in the city public schools, and is lowest in the suburban public high schools; (*c*) the percentage of smokers is highest among students who do not participate in any school activities; (*d*) the percentage of smokers in this young age group is inversely related to the educational level of the parents; (*e*) the percentage of smokers is higher among those students who are behind their age-equals scholastically, thus confirming the findings of Earp (1936) and of Lynn (1948).

Emotional status and cigarette smoking. As part of his study of 903 smokers and 903 nonsmokers, Lilienfeld (1959) utilized a 31-item questionnaire made up from a list developed by Stauffer et al. The 31 items selected were those which differentiated a "normal" from a "neurotic" group. Examination of the 31 items used by Lilienfeld (1959, pp. 264–268) makes clear that this questionnaire utilized items very similar to those contained in the MAS, CMI, and the Saslow Psychosomatic Screening Inventory (see section below).

Lilienfeld's findings (p. 269) were that "the responses by cigarette smokers on the questions concerning emotional status were consistently more 'neurotic' than those of nonsmokers." Nineteen of the 31 items reached statistical significance: 11 at the .001 level of confidence, 4 at the .01 level, 1 at the .02 level, and 3 at the .05 level, all in the direction stated. The remaining 12 items were nondiscriminating.

An additional finding (p. 278) was that the smokers had had a significantly greater number of hospitalizations ($p < .001$). The author suggested that while the excess number of hospitalizations may reflect neuroticism, they also may reflect other reported associations between disease and cigarette smoking. Lilienfeld (p. 276) reports that an analysis of the reasons for these hospitalizations is now in progress.

These findings of Lilienfeld, while in agreement with the findings of Lawton and Phillips, and our own, are important for several additional reasons: (*a*) the sample was large ($N = 1,806$) and was an adequate representation of the entire normal adult population of a large city; (*b*) the sample contained two groups matched on four important variables (age, sex, race, and social status); (*c*) there was a reliability check on the findings. Unlike the two other studies which used selected subgroups in the population (university undergraduates, student nurses, psychiatric patients from a large medical center, and VA hospital patients), Lilienfeld's study dealt with a noninstitutional sample of adult normals which was more representative of the total sample of adults in the United States than were the other specialized samples. Yet the fact that his findihgs and those of the other investigators agree is striking.

At this point a comment is in order about a potential shortcoming of these three questionnaire studies

and several of the others here being reviewed: The differences between smokers and nonsmokers were based on *self-report*. While it is probably true that smokers and nonsmokers have similar "test-taking attitudes," there is always the possibility that the "response-set" of the two groups differs in such a way that smokers have a lower threshold for admitting "neurotic" items than do nonsmokers. No study suggesting such a possibility is known to us. On the other hand, there are sufficient (neurotic) indices not based on self-response questionnaires which significantly differentiated the two groups (e.g., Lawton and Phillips' finding of a greater incidence of Heavy Smokers in their psychiatric sample in contrast to their nonpsychiatric control sample) that, in the absence of data suggesting otherwise, it can be presumed that smokers and nonsmokers have similar test-taking attitudes, and that the obtained differences reflect significant differences between the two groups. Even if it were shown that test-taking attitudes *were* different, this fact, by itself, would still be an important psychological difference between smokers and nonsmokers.

Psychosomtic screening inventory scores and cigarette smoking. Table 4 contains for the three populations studied by us mean scores for smokers and nonsmokers on the Saslow Psychosomatic Screening Inventory (Gleser & Ulett, 1952; Saslow, Counts, & DuBois, 1951). This test requires the respondent to check from a list of 23 symptoms (dealing with bodily and mood dysfunction) the particular symptoms he or she experiences in everyday anxiety and anger situations.

In all three populations studied, the smokers report a greater number of psychosomatic symptoms than do nonsmokers. While these differences reach statistical significance only in the student nurse group, the trend is obviously in the same direction with the two other samples.

Coffee and alcohol consumption and cigarette smoking. In our own three groups, coffee intake was scored as number of cups consumed per day, while alcohol intake was estimated on a weekly basis by a crude scale that gave differing weights to beer, wine, and whiskey. A score of 0 was given to nondrinkers while a score of 5 was assigned to 21 or more glasses of beer per week, or 21 or more glasses of wine, or 21 or more single shots of whiskey.

The results shown in Table 4 are all in the same direction, and reach statistical significance ($p < .001$) in two of the four comparisons for coffee consumption. Thus, smokers consume more coffee. For liquor, again the trend is in the same direction for all groups, and the differences reach statistical significance ($p < .001$ and .05) in two of the four comparisons. Smokers also consume more alcohol.

Both these findings ($p < .01$) have been reported independently by McArthur et al. (1958, p. 269) and Heath (1958, p. 385).

An interesting observation in the literature on the possible interaction between coffee and cigarette intake is that by Troemel, Davis, and Hendley (1951). Studying a phenomenon (dark adaptation), long of interest to experimental psychologists, these authors report an increased speed of dark adaptation after inhaling cigarette smoke, which could be counteracted by the simultaneous intake of caffeine in a small dose but not by a larger dose of caffeine.

Little systematic research has appeared describing the relationship between food intake and smoking. However, smokers occasionally re-

port that smoking helps them keep down their weight. Brozek and Keys (1957), in a study indirectly bearing on this point, have reported that over a 5-year period, smokers who discontinue smoking for a 2-year period show a statistically significant ($p<.001$) weight gain compared either to themselves or to a control group of smokers who continue to smoke. The smokers who stopped smoking showed a weight gain of 3.73 kg. (8.2 lb.) relative to their own previous weight, while the continuing smokers showed a weight loss of .50 kg. (not significant). Hammond and Percy (1958) recently have confirmed this weight increase when smoking is discontinued. These authors found that, of 333 ex-smokers, 246 (73.9%) said they gained weight when they stopped smoking.

Long-Term Studies of Smokers

We were able to find only one such study: a 20-year longitudinal study of 252 smokers and nonsmokers from Harvard College. These were all participants in the well-known Study of Adult Development (Grant Study), a long-term study of selected "normal" college men who were studied during their undergraduate years (1939–1942) and who have been followed by interview and questionnaire from that time to the present—a period of 20 years. The smoking data were incidental to the main objectives of the study, but have recently been reported by McArthur et al. (1958) and Heath (1958). Their reports contain a wealth of anecdotal and statistical comparisons which are not easily summarized and for which their publications should be consulted.

Of the 252 subjects 61 (24.2%) were nonsmokers, 95 (37.7%) were moderate smokers, and 96 (38.1%) were heavier smokers. During the 20-year period heavy smokers and nonsmokers showed a number of contrasting characteristics which are summarized in our Table 5 taken from Heath (1958, p. 387).

The results of this study are im-

TABLE 5

Contrasting Characteristics of Nonsmokers and Heavier Smokers[a]

Nonsmokers	Heavier Smokers
Bland affect	Cultural
Inarticulate	Lack of purpose and values
Well-integrated	Less well-integrated
Physical sciences	Practical organizing
Most stable personality	Less stable personality
Major in college: natural sciences	Major in college: social studies, arts and letters
Careers: chemistry, physics	Careers: social relations, education
Like: science research worker	Like: judge
Dislike: sales manager	Dislike: science research worker
Psychotype: cerebrotonia	Psychotype: viscerotonia
Answer questionnaires promptly	Delay answering questionnaires
Armed Service: Navy; well-adjusted, noncombat duty	Armed Service: Army; less well-adjusted, combat duty
Respiratory rate: slow	Respiratory rate: rapid
Sighs and swallows: diminished	Sighs and swallows: increased
Reflexes: increased	Reflexes: decreased
Less alcohol and coffee	More alcohol and coffee

[a] From Heath (1958, p. 387).

portant because, as a 20-year longitudinal study, it contains the only published observations of their kind. The findings point to many psychological variables which deserve further study. However, since many hundreds of variables were studied by the several Harvard investigators, and since apparently so few variables (those shown in Table 5 and a few others) reached statistical significance, there is always left open the question whether or not the reported findings are due to chance phenomena. McArthur et al. are aware of this possibility (1958, pp. 273–274) and offer some cross-validational findings from two classes of recent Harvard undergraduates. The bulk of the findings, while very suggestive and stimulating, needs confirmation, however.

Factors Influencing the Initiation of Smoking

Despite the likely importance of this subject, little systematic study has been made of the possibility that the factors associated with the initiation of smoking may not be the same as those associated with its continuation. In our own opinion, there is considerable a priori reason to believe that the factors which motivate people (especially teenagers) to start to smoke are probably very different from those factors which help perpetuate the habit once smoking has gone on for any length of time.[3]

The earlier described study by Horn et al. (1959) of 22,000 high school students points to factors associated with the initiation of smoking (parental smoking habit, participation in school activities, etc.) which could hardly be thought to have equal importance for continuation of smoking in later years. This age group provides many opportunities to study both peer and parental influences associated with teenagers' beginning to smoke. The recently published results undoubtedly will prove interesting and provocative to other investigators. Except for the few indirect hints arising from the 20-year follow-up study of Heath (1958) and McArthur et al. (1958), little is known about why Americans continue smoking once they have started.

A second study has been carried out by Phanishayi (1951). Although the number of subjects studied was small (48 male college graduate and postgraduate students), and the country (India) not our own, he was able to gather suggestive data on self-reported reasons why these college men (mean age 25) began to smoke and also, once having started, why they continued to smoke. Phanishayi prepared a 48-item questionnaire on which each student was required to check for 24 items those reasons, and only those reasons, which were associated with his beginning to smoke and, for the remaining 24 items, only those reasons why he continued to smoke. Some of the items in the two lists were identical. The two parts of the questionnaire thus had certain elements peculiar to each smoking stage and certain others common to both. At the end of the 24 statements in both Parts I and II, a provision was made for the subject to give other causes if he had any. In addition, each subject was asked to rank-order his own choices as to the three most important reasons why he began and why he continued to smoke. The 48 subjects were told the purpose of the investigation and

[3] To examine this hypothesis, personnel and facilities of our department were made available to Horn of the American Cancer Society.

were requested to try to be true to facts as far as possible. The results of this study were as follows.

Initiation of smoking: the reasons checked for the beginning of the smoking habit are diverse and varied, although the most frequently stated reason is a curiosity as to the nature of pleasure afforded by smoking. The most frequent reasons for beginning and the percentage of the students listing each were as follows:

I wanted to see what sort of pleasure I would get out of it (75%).
I thought there was nothing wrong in doing so (52%).
I thought it would help me sit up during nights for study (50%).
I thought there must be something attractive about it because so many people do so (48%).
I thought that a trial would cost nothing (48%).

Two of the reasons with the lowest frequencies were:

I was tempted by the advertisements (4%).
I wanted to rebel against the authority of my parents and smoking is one of the ways of doing so (4%).

Continuation of smoking: the reasons checked why smoking is continued seem to have a strong psychological basis, especially involving reduction of anxiety, tension, and loneliness. The most important stated reasons why these 48 students continued to smoke and the percentages were as follows:

It serves as a companion when I am alone (75%).
It warms me up when I am cold (71%).
It helps to forget my worries and anxieties (60%).
It facilitates thinking and gives inspiration (52%).
It enables me to acquire new friends easily (44%).

The lowest frequency reason given for continuing to smoke was:

To give up would be a submission to orthodoxy (4%).

Despite the few subjects studied, Phanishayi's results suggest (1951, p. 36) that

the causes which prompt a person to take to the habit (curiosity, especially) are different from those which are responsible for its continuation (tension-reduction capabilities, especially).

There is every reason to expect that the several expected reports from the sample of 22,000 high school students of Horn et al. will yield results for beginning the habit similar to Phanishayi's, in addition to others which Horn's different methodology may unearth. As a matter of fact, unpublished findings in the longer report of Schubert (1959) of 226 northeastern United States college students indicates that the reasons given by these students for beginning to smoke were not too unlike those of Phanishayi's sample. Schubert (1959) also found that for both his male and female subjects three scales of the MMPI (*Ma*, *Hy*, and *Pd*) differentiated smokers and nonsmokers.

As to the reason for *excessive* smoking once the habit has begun, Bergler (1946, 1953), on the basis of five cases of compulsive smokers who underwent psychoanalysis, suggests

In all cases, a similar neurotic substructure was found: These patients represented a specific type of masochist who *unconsciously* wanted constantly to be *refused*. To counteract the inner reproach stemming from their inner conscience, which objected to the wish to be refused, they "proved" that they wanted the opposite—to *get*. The outward sign was getting a cigarette, an oral "pacifier," reminding the individual of his first reassurance in life (1946, p. 320).

Factors Influencing the Termination of Smoking

How many people quit smoking once they start is not yet well-known. Recent studies (Haenszel, et al., 1956, p. 24; Hammond & Percy, 1958, p.

2956) place the number at under 20%. One of the best studies is the recent one by Hammond and Percy (1958).

Hammond and Percy study. Their population and results were as follows:

> Of 3,560 out of 5,992 men (selected from telephone directories all over the country) who filled out a questionnaire on smoking habits, 2,498 (70.2 per cent) stated that they smoked cigarettes regularly or had done so in the past. Of these 2,498 men, 472 (18.9 per cent) stated that they no longer smoked cigarettes or tobacco in any form. A questionnaire asking why they had stopped smoking was sent to the 472 ex-smokers, and 333 (70.6 per cent) replied.
>
> Only 6.3 per cent of the 333 ex-smokers said that they gave up the habit because of reports linking smoking to lung cancer, and an additional 2.4 per cent said that they gave it up because of reports that smoking has a bad effect on health in general. In other words, only 1.6 per cent of men with a history of regular cigarette smoking said that they gave up the habit because of reports relating cigarette smoking to lung cancer or other diseases. Some condition apparently made worse by smoking was given as a reason for stopping by 208 (62.5 per cent) of the 333 ex-smokers. Coughing was the most frequently mentioned reason for giving up the habit.
>
> Some improvement, such as less coughing, less shortness of breath, etc. was noted by 272 (81.7 per cent) of the men as an apparent result of giving up smoking. Of the ex-smokers, 246 (73.9 per cent) said that they gained weight when they stopped smoking.
>
> . . . It is obvious that only a very small percentage of cigarette smokers have given up the habit consciously and admittedly because of reports linking cigarette smoking to lung cancer and other serious diseases. Even if the reported link with lung cancer was a contributory factor in several times as many cases as recorded on these questionnaires, it still was relatively unimportant in terms of inducing cigarette smokers to stop. On the other hand, reports on lung cancer and longevity may well be a major factor in the remarkable increase in popularity of filter-tip cigarettes which are advertised as having "less tar and nicotine" (p. 2959).

That the findings of Hammond and Percy that the many reports linking lung cancer and cigarette smoking have apparently induced only a few smokers to give up the habit are probably an accurate estimate of the facts for the general public is dramatically attested to in a second study.

Lawton and Goldman study. These investigators (1958) conducted a survey of 72 internationally renowned *lung cancer scientists* who attended the American Cancer Society sponsored conference on lung cancer at Virginia Beach in 1957. Their control group consisted of 72 psychologist-scientists from the Division of Experimental Psychology of the American Psychological Association. Both groups were matched for age and sex, and roughly for scientific nature of their interests.

Interestingly, the two groups did differ significantly in (*a*) the percentage of current cigarette smokers ($p<.01$, one-third fewer current smokers among the lung cancer scientists), and (*b*) past incidence of cigarette smoking ($p<.05$, with one-third more lung cancer scientists having never smoked). This difference in current and life-long smoking pattern, with more nonsmokers among the lung cancer scientists in contrast to the the behavioral scientists, would appear to be an independent confirmation of Heath's (1958) finding in the 20-year follow-up study of college major and career choices of Harvard smokers and nonsmokers described earlier (see Table 5).

Lawton and Goldman's next finding was not surprising in view of the different areas of specialization involved (and thus presumed firsthand familiarity with the scientific literature in each of the two fields): 83.3% of the lung cancer scientists versus only 63.6% of the psychologist-scientists felt that cigarette smoking is a cause of lung cancer ($p<.01$). When

the number of smokers in both groups was held constant, this significant difference still existed.

Despite this greater knowledge and/or conviction in the lung cancer scientist group regarding the health hazard brought on by cigarette smoking, Lawton and Goldman's next finding was that there was not a significantly greater number of lung cancer scientists who had quit smoking in the period 1952–1957; nor did a significantly greater percentage of them express a dissatisfaction with their smoking habit (25% in one group and 27% in the other expressed dissatisfaction that they were cigarette smokers).

A final part of their study showed that expressed attitude toward lung cancer causation did have a significant effect upon the smoking behavior of the psychologist sample of smokers: smoker-psychologists who felt that smoking is a cause of lung cancer tended to have a lower current incidence of smoking ($p < .02$), attempted to cut down the amount of their daily consumption significantly more ($p < .01$), had a greater number of (unsuccessful) attempts to stop smoking ($p < .05$), and expressed more dissatisfaction with their current smoking habit ($p < .02$). The fact that there were so few (3 out of 70) lung cancer scientists who felt that smoking does not cause lung cancer made a similar analysis of these attitudes for this group not meaningful. However, the fact that, relative to the psychologists, there was not a significantly greater number of quitters in the lung cancer scientist group, with their greater expressed conviction of the health hazard, again emphasizes the importance of noncognitive elements in smoking behavior, and highlights the fact that even for health experts the health hazard per se is not a sufficient deterrent to cigarette smoking. Thus, one is not surprised that Hammond and Percy (1958) found that only 6.3% of ex-smokers said they quit smoking because of the lung cancer reports. Whatever factors motivate continuance of smoking are thus seen to be more potent than scientifically accepted probability statistics showing a greater health risk.

Study of Harvard men. McArthur et al. (1958, pp. 272–273), reasoning from a very small sample of those smokers in their sample who could stop and those smokers who could not break the habit, found some slight suggestions that early breast-feeding habits and later personality integration seem to be related to ability or inability to quit smoking. However, these findings and those of Bergler (1946, 1953) on the reasons for excessive smoking probably will have to be investigated by other measures of "personality integration" in addition to the clinical interview and the Rorschach inkblot technique.

Discussion

Table 6 is a summary of the characteristics which do, and those which do not, differentiate smokers from nonsmokers. It is clear from the studies just reviewed that our knowledge of the personality and psychosocial characteristics of smokers and nonsmokers is only in its beginnings. Nevertheless the studies do suggest some answers to pertinent questions which have been raised in this general area.

First, the question of a "*smoker's personality.*" Inspection of Tables 1 through 5, plus consideration of the other studies reviewed makes clear that while smokers do differ from nonsmokers in a variety of character-

TABLE 6

SMOKERS AND NONSMOKERS: SUMMARY OF DATA REVIEWED

Differentiating Characteristics	Nondifferentiating Characteristics
Age	Race, United States white versus United States nonwhite
Sex	Hollingshead Index of Socioeconomic Status
Foreign-born parents	Education, highest grade attained
Current marital status; marital history	IQ
Occupation, including unemployment	Mental and motor functioning, immediate effects upon
Military service, current or history of	
Frequency of job changes	
Urban-rural residence	
Residence, United States coastal versus noncoastal	
Income	
Social class, including social mobility	
College graduation history	
Delicate motor coordination	
Characteristics from 20-year study summarized in Table 5	
Body sway suggestibility	
Driving accidents	
Participation in sports	
Taylor anxiety score	
Psychological tension level (CMI)	
Psychiatric versus medical status	
Psychosomatic symptoms, number of	
Emotional status ("neuroticism")	
Extraversion-introversion	
Number of hospitalizations	
Coffee consumption	
Alcohol consumption	
Weight increase	
Dark adaptation	
Cancer-scientist versus psychologist-scientist	
Smoking habits of parents	
Educational level of parents	
Own education-age concordance	
Parochial versus public school attendance	
Participation in school activities	

istics, none of the studies has shown a single variable which is found exclusively in one group and is completely absent in the other. While this is true for *all* of the variables summarized in the appropriate column in Table 6, it is especially true for the variables measuring personality characteristics. Thus, while Taylor anxiety score, and Cornell Medical (psychiatric) Index, etc. significantly differentiate smokers from nonsmokers, the mean differences are not large. Examination of the means, standard deviations, ranges, percentages, etc. of the various published studies makes clear that while group trends suggest the smoker to be more "neurotic," *on the average*, there are still many individual smokers with neuroticism, or anxiety, etc. scores lower than those of many nonsmokers, and vice versa. Thus a clear-cut smoker's personality has not emerged from the results so far published in the literature. This is not surprising

when it is remembered that approximately 60 million Americans over the age of 18 smoke. It is hard to believe that they would share in common one personality "type." This is not to imply, however, that the various psychological dimensions along which smokers have been shown, as a group, to differ from nonsmokers may not suggest an important single process, or processes, underlying these various demonstrated differences. Further research may indeed so systematize the disparate findings.

While the evidence for a smoker's personality is not strong, it is possible to think that there may be in certain individuals a biological or *genetic predisposition* to a strong desire to smoke (as well as to lung cancer). To establish that cigarette smoking is genetically determined would require studies yielding more definitive data than are as yet available.[4] Nevertheless, Fisher (1958) believes this genetic hypothesis has some merit on the basis of a reported German study of monozygotic and dizygotic twins—in which the smoking behavior of 51 pairs of monozygotic twins is reported to be more nearly alike than that of 31 pairs of dizygotic twins. Fisher writes:

> The data so far assembled relate to 51 monozygotic and 31 dizygotic pairs, from Tübingen, Frankfurt and Berlin. Of the first, thirty-three pairs are wholly alike qualitatively, namely, nine pairs both non-smokers, twenty-two pairs both cigarette smokers and two pairs both cigar smokers. Six pairs, though closely alike, show some differences in the record, as in a pair of whom one smokes cigars only, whereas the other smokes cigars and sometimes a pipe. Twelve pairs, less than one-quarter of the whole, show distinct differences, such as a cigarette smoker and a non-smoker, or a cigar smoker and a cigarette smoker.
>
> By contrast, of the dizygotic pairs only eleven can be classed as wholly alike, while sixteen out of thirty-one are distinctly different, this being 51 per cent against 24 per cent among the monozygotics.
>
> The data can be rearranged in several ways according to the extent to which attention is given to minor variations in the smoking habit. In all cases, however, the monozygotic twins show closer similarity and fewer divergencies than the dizygotic.
>
> There can therefore be little doubt that the genotype exercises a considerable influence on smoking, and on the particular habit of smoking adopted, and that a study of twins on a comparatively small scale is competent to demonstrate the rather considerable differences which must exist between the different groups who classify themselves as non-smokers, or the different classes of smokers (p. 108).

Hammond (1958), on the other hand, believes this genetic hypothesis untenable:

> It has been suggested that there may be some hereditary factor which results in both lung cancer and a strong desire to smoke cigarettes. This is an ingenious idea. However, if it is true, one must assume that (since cigarette smoking has increased so considerably in the past half century) a genetic factor of this sort appeared and became widely spread throughout the populations of many countries during the last fifty years. This seems a bit unlikely. Anyone with a good imagination can think of other such conceivable mutual causation hypotheses, but no one has yet presented any evidence in support of them. . . . Other evidence, such as the sex difference, the urban-rural difference, the time trends in lung cancer death rates and cigarette consumption is consistent with the causal hypothesis. . . . With so much evidence all pointing in the same directon and no evidence pointing in any other direction, I can only arrive at one conclusion. In my opinion, cigarette smoking is a major factor in the causation of lung cancer (p. 350).

[4] After this review went to press, Eysenck, Tarrant, Woolf, and England (1960), in a study of 2,360 British male subjects, reported finding that cigarette smokers are more extroverted than nonsmokers. Reasoning from Eysenck's earlier work on a suggested genetic basis for his extraversion factor, they concluded, therefore, that genotypic differences exist between smokers and nonsmokers, and between cigarette smokers and pipe smokers. They suggest a number of further studies to establish this hypothesis even more soundly.

It is clear from his remarks that Hammond is as skeptical of the evidence suggesting a hereditary predisposition to the desire to smoke as he is to an alleged hereditary predisposition to lung cancer.

However, in fairness to the genetic hypothesis, it should be pointed out that the possibility exists of a constant frequency of genetic predisposition combined with a progressive increase and availability of cigarettes during the past 50 years. In short, an interacting, ecological relationship could exist between genetic make-up and culturally determined economic conditions (mass production, marketing, etc.).

A third question that deserves consideration is the possibility that smoking behavior is determined by *multiple factors* rather than by a single factor (personality or genotype). Table 6 suggests that for any given individual, *his* smoking behavior may be socioculturally determined, age-sex linked, related to occupation, and/or associated with a variety of personality and other behavioral characteristics. The present meagre available evidence does not permit the determination of the relevant weight of these pertinent factors singly. Nor are there studies which show the multiple correlation with smoking across individuals, or within a single individual, of the various characteristics shown in Table 6.

In addition to these uncertainties, it should be added that none of the studies here reviewed provides an answer to the "cause and effect" question in the relationship between psychological characteristics and smoking behavior. It will be clear to the reader that the studies just reviewed showing, for example, slightly more "anxiety" or higher "neurotic indices" in smokers compared to nonsmokers were merely studies of *a relationship* between smoking and these personality variables, since in no way were the studies designed to investigate cause and effect. There could be a complex mutually causative relationship between smoking and the various characteristics shown in Table 6 (as well as smoking and lung cancer and/or coronary heart disease). As Hammond (1958) points out, at this stage of our knowledge, for example,

> there is as much reason to suppose that cigarette smoking causes nervous tension as to believe that nervous tension causes cigarette smoking. Perhaps they are mutually causative as in an autocatalytic type of reaction (p. 352).

A similar statement regarding cause and effect could be made about each of the many other characteristics related to smoking shown in Table 6.

Longitudinal studies of youngsters before and after they have begun to smoke might provide some leads to a partial answer to the "cause versus effect" question. One such study might employ an anxiety questionnaire and could provide information as to whether nonsmoking youngsters with higher anxiety scores begin to or do not begin to smoke in greater percentages at a given age than do youngsters with lower anxiety scores. In addition, readministration of such an anxiety test several years later would permit study of the question whether or not youngsters who take up smoking later earn higher anxiety scores than control youngsters who have never taken up smoking; or earn higher anxiety scores relative to their own presmoking anxiety scores. Numerous other designs could supply additional information regarding cause, effect, and mutual interaction.

In summary one can suggest that, while this review and the data sum-

marized in Table 6 make clear that smokers and nonsmokers do, in fact, differ on a number of psychological, personal, social, and behavioral characteristics, it is equally clear that research in this area (including appropriate selection and sizes of samples, controls, etc.) has just begun. The number of recent psychological studies of smokers and nonsmokers suggests that publications in the next few years may reflect an even greater interest among behavioral scientists in this very common form of human behavior.

REFERENCES

ALLEN, BERNADENE V. An investigation of the relationship between smoking and personality. Unpublished MA thesis, Univer. Portland, 1958.

BACKETT, E. M. Advances in preventive medicine. *Practitioner*, 1958, **181**, 494–502.

BERGLER, E. Psychopathology of compulsive smoking. *Psychiat. Quart.*, 1946, **20**, 297–321.

BERGLER, E. Smoking and its infantile precursors. *Int. J. Sexol.*, 1953, **6**, 214–220.

BRODMAN, K., ERDMANN, A. J., & WOLFF, H. G. *The Cornell Medical Index manual.* New York: Cornell Univer. Medical Coll., 1949.

BROZEK, J., & KEYS, A. Changes of body weight in normal men who stop smoking cigarettes. *Science*, 1957, **125**, 1203.

EARP, J. R. Cited by A. G. Ingalls in If you smoke. *Scient. Amer.*, 1936, **154**, 310–313.

EYSENCK, H. J., TARRANT, MOLLIE, WOOLF, MYRA, & ENGLAND, L. Smoking and personality. *Brit. med. J.*, 1960, **1**, 1456–1460.

FISHER, R. A. Lung cancer and cigarettes. *Nature*, 1958, **182**, 108.

GLESER, GOLDINE, & ULETT, G. A. The Saslow Screening Test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, **8**, 279–283.

HAENSZEL, W., SHIMKIN, M. B., & MILLER, H. P. Tobacco smoking patterns in the United States. *Pub. hlth. Monogr.*, 1956, No. 45. (Public Health Service Publication, 1956, No. 463)

HAMMOND, E. C. Smoking and death rates: A riddle in cause and effect. *Amer. Scientist*, 1958, **46**, 331–354.

HAMMOND, E. C., & HORN, D. The relationship between human smoking habits and death rates: A follow-up study of 187,766 men. *J. Amer. Med. Ass.*, 1954, **155**, 1316–1328.

HAMMOND, E. C., & HORN, D. Smoking and death rates: Report on 44 months of follow-up of 187,783 men. *J. Amer. Med. Ass.*, 1958, **166**, 1159–1172, 1294–1308.

HAMMOND, E. C., & PERCY, CONSTANCE. Ex-smokers. *NY State J. Med.*, 1958, Sept., 2956–2959.

HEATH, C. W. Differences between smokers and non-smokers. *AMA Arch. intern. Med.*, 1958, **101**, 377–388.

HINKLE, L. E., & WOLFF, H. G. Health and the social environment: Experimental investigations. In A. H. Leighton, J. A. Clausen, & R. N. Wilson (Eds.), *Explorations in social psychiatry.* New York: Basic Books, 1957. Pp. 105–137.

HOLLINGSHEAD, A., & REDLICH, F. *Social class and mental illness.* New York: Wiley, 1958.

HORN, D., COURTS, F. A., TAYLOR, R. M., & SOLOMON, E. S. Cigarette smoking among high school students. *Amer. J. publ. Hlth.*, 1959, **49**, 1497–1511.

HULL, C. L. The effects of tobacco smoking on mental and motor efficiency. *Psychol. Monogr.*, 1924, **33**(3, Whole No. 150).

KALLNER, GERTRUDE. Smoking habits of the population. *Statist. Bull. Israel*, in press. (English translation of a mimeographed report, Central Bureau of Statistics, State of Israel, November 1958.)

KIRCHOFF, HELEN, & RIGDON, R. H. Smoking habits of college students in Texas. *Tex. Rep. Biol. Med.*, 1954, **12**, 292–299.

LAWTON, M. P., & GOLDMAN, A. E. Cigarette smoking and attitude toward the etiology of lung cancer. *Amer. Psychologist*, 1958, **13**, 342. (Abstract)

LAWTON, M. P., & PHILLIPS, R. W. The relationship between excessive cigarette smoking and psychological tension. *Amer. J. med. Sci.*, 1956, **232**, 397–402.

LILIENFELD, A. M. Emotional and other selected characteristics of cigarette smokers and nonsmokers as related to epidemiological studies of lung cancer and other diseases. *J. Nat. Cancer Inst.*, 1959, **22**, 259–282.

LITTLE, C. C. The public and smoking: Fear or calm deliberation. *Atlantic*, 1957, **200**,

74–76. (Reprinted: *Ca: Bull. Cancer Progress*, 1958, **8**, 49–52.)

Lynn, R. M. A study of smokers and nonsmokers as related to achievement and various personal characteristics. *U. NC Rec. Res. Progress*, 1948, No. 464, 164. (Grad. Sch. Series No. 56) (Abstract)

McArthur, C., Waldron, Ellen, & Dickinson, J. The psychology of smoking. *J. abnorm. soc. Psychol.*, 1958, **56**, 267–275.

Matarazzo, J. D., Guze, S., & Matarazzo, Ruth G. An approach to the validity of the Taylor anxiety scale: Scores of medical and psychiatric patients. *J. abnorm. soc. Psychol.*, 1955, **51**, 276–280.

Matarazzo, Ruth G., Matarazzo, J. D., Saslow, G., & Phillips, Jeanne S. Psychological test and organismic correlates of interview interaction patterns. *J. abnorm. soc. Psychol.*, 1958, **56**, 329–338.

Northrup, E. *Science looks at smoking: A new inquiry into the effects of smoking on your health.* New York: Coward-McCann, 1957.

Phanishayi, R. A. Causes of smoking. *J. Educ. Psychol., Baroda*, 1951, **9**, 29–37.

Raven, R. W. Smoking habits of schoolboys. *Lancet*, 1957, June, 1139–1141.

Russ, S. *Smoking and its effects.* New York: Macmillan, 1956.

Rutstein, D. D. An open letter to Dr. Clarence Cook Little. *Atlantic*, 1957, **200**, 41–43. (Reprinted: *Ca: Bull. Cancer Progress*, 1958, **8**, 46–48.)

Sackrin, S. M., & Conover, A. G. Tobacco smoking in the U. S. in relation to income. *USDA Market. res. Rep.*, 1957, No. 189.

Saslow, G., Counts, R. M., & Du Bois, P. H. Evaluation of a new psychiatric screening test. *Psychosom. Med.*, 1951, **13**, 242–253.

Schubert, D. S. P. Personality implications of cigarette smoking among college students. *J. consult. Psychol.*, 1959, **23**, 376.

Study Group on Smoking and Health. Smoking and health: Joint report of the Study Group on Smoking and Health. *Science*, 1957, **125**, 1129–1133.

Super, D. E. *Appraising vocational fitness by means of psychological tests.* New York: Harper, 1949.

Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

Troemel, R. G., Davis, R. T., & Hendley, C. D. Dark adaptation as a function of caffeine and nicotine administration. *Proc. S. Dak. Acad. Sci.*, 1951, **30**, 79–84.

Vallance, T. R. Suggestibility of smokers and nonsmokers. *Psychol. Rec.*, 1940–45, **4–5**, 138–144.

(Received June 17, 1959)

Psychological Bulletin
1960, Vol. 57, No. 6, 514–532

THE STRUCTURE OF ABILITIES AT THE PRESCHOOL AGES: HYPOTHESIZED DOMAINS[1]

C. E. MEYERS AND HARVEY F. DINGMAN
University of Southern California *Pacific State Hospital*

This paper discusses the possibilities for factorial descriptions of the abilities of the infant and young child. It brings forth the practical and theoretical need for test reconstruction and shows the potentials for doing so which reside in the current repertory of test materials. Finally the paper hypothesizes a series of factors which may ultimately be identified at the late preschool level, with some speculations regarding the early appearance of the hypothesized factors.

Characteristics of Ability Examinations for Infants and Young Children

Most of the instruments in current use for appraisal of the abilities at the preschool ages were constructed after the prototypes of Binet and Gesell. In both instances these pioneers faced a need to describe a child in direct terms and for practical usage. They took the easiest approach possible. The examined child was tried out on the behaviors of his own and other age groups, and was then described in terms of the age norms of the test items he passed.

The test construction and reporting of results so developed are illustrated in the construction and use of the popular instruments in contemporary clinical practice. Today's age-scale construction also employs the two technical criteria for selective retention of tried-out test items, age-progression and internal consistency. Inasmuch as the "common" behavior of young children changes with growth, the employment of these technical criteria sharpens any age-to-age difference in the abilities sampled at different levels of the examination, and narrows the spectrum of abilities sampled at any one level.

Table 1 lists a series of standardized examining instruments available for use at the various preschool ages. Most of the instruments of Categories A through D, which are the clinical tests of most concern here, were constructed on Binet-type principles and procedures, and yield an "age score" as a test result. The tests apparently satisfy enough of the clinical need to describe a child as he is now, in terms of age norms.

It can be shown that the instruments are not the best conceivable ones for certain other purposes. One such purpose is that of clinical prediction. Over a dozen studies have demonstrated that baby tests and early preschool tests do not predict later intelligence very well (most recently, Cavanaugh, Cohen, Dunphy, Ringwall, & Goldberg, 1957; Wittenborn, 1956). It is also shown that the best items for predicting future status may be poor items on a criterion of internal consistency at the level where they are placed (Nelson & Richards,

[1] Supported in part by the National Institute of Mental Health Grant 3M-9130: Population Movement of Mental Defectives and Related Physical, Behavioral, Social, and Cultural Factors. Pacific State Hospital, Pomona, California.

TABLE 1

ILLUSTRATIVE CURRENT ABILITY MEASURES WITH NOTATIONS ON ITEM SELECTION AND AGE-SCORING

Type of Instrument and Stated or Implied Scope	Preschool Ages Tested	Bases for Item Selection	Kind of Score Yielded
A. General Behavioral Development			
Gesell & Amatruda (1947) "Gesell schedules"	1 month up	Age-normal behavior	DA (by median success level)
Griffiths (1954) Abilities of babies	1–24 months	Age-normal behavior; "G" broadly conceived	Cumulated MA
B. General Intelligence			
Cattell (1940) Infant Intelligence Scale	3–24 months	Age-normal behavior; some selection by face validity for intelligence	Cumulated MA
Shotwell, Dingman, & Tarjan (1957) Number concept test	3 years up	Quantitative tasks, progression with MA	Cumulated MA
Terman & Merrill (1937) Stanford-Binet	2 years up	r with 1916 SB; face validity, age progression, internal consistency	Cumulated MA
Valentine (1950) Intelligence tests for children	2 years up	Age-normal; face validity	Cumulated MA
Ammons & Ammons (1948) Picture vocabulary	2 years up	Age-suitable recognition vocabulary, r with SB	MA from points; centiles
C. General Intelligence by Nonlanguage, Culture-Free Means			
Arthur (1947) Point Scale	4½ years up	Face validity for intelligence; r with other tests	MA from points
Burgemeister, Blum, & Lorge (1954) Columbia Mental Maturity	3 years up	Nonverbal, ease of resp., suitable to age, face, r with other tests	MA from points
Leiter (1948) International Performance Scale	2 years up	Age-suitable culture-fair, face validity for G	Cumulated MA
D. Purposes Other than to Measure Intelligence			
Bayley (1935) California Motor Scale	1–50 months	Age-normal motor	Points; sigma for age
Doll (1953) Social Maturity Scale	1 year up	Age-normal, self-help, culture-required conduct	Social age by points
Sloan (1955) Lincoln-Oseretsky Motor Tests	4 years up	Motor, selected from Oseretsky by objectivity, age progression	Points to centiles by age and sex

TABLE 1—(*Continued*)

Type of Instrument and Stated or Implied Scope	Preschool Ages Tested	Bases for Item Selection	Kind of Score Yielded
E. School Readiness Examinations			
Hildreth & Griffiths (1949) Metropolitan Readiness Tests	kinderg.-primary	Face validity via analysis of primary reading; r with later reading	Centile of points for Grade 1
Lee & Clark (1951) Reading Readiness Test	kinderg.-primary	Face validity by analysis of primary reading	Centile of points for Grade 1
Monroe (1935) Reading aptitude	kinderg.-primary	Analysis of primary reading; r with later reading	Centile of points for ages
F. Batteries for Testing Differentiated Abilities			
Thurstone & Thurstone (1953) Primary mental abilities	5–7 years	Factor analysis	Factor points to factor MAs; total MA
Sievers (1955) Illinois language tests	2–6 years	Osgood psycholinguistics model	Subtest and total point norms by age (11 tests)

1938, 1939). All such studies (except those of Gesell) were preoccupied with the prediction of future "intelligence," the criterion variable being a later Binet or other IQ. Little has been done by psychologists on the prognostic value of early testing for other variables of social and theoretical interest. Will the motor subscale in Gesell and Amatruda (1947) tell of later athletic prowess or the age of ambulation in handicapped children? Do the personal-social items in Griffiths (1954) predict leadership in the fifth grade? Surprisingly little is available. Neilon (1948) reported good consistency in general behavior descriptions between infancy and adolescence in Shirley's (1933) famous subjects, and provides a good review of available literature.

Gesell (e.g., Gesell, Castner, Thompson, & Amatruda, 1939; a summary in 1954) found consistency within broad categories of diagnosis, such as mentally deficient or palsied, and some intriguing instances of temperamental consistency in normals as well. Escalona (1950) and Gallagher (1953) demonstrated that prediction from infant testing gave better results if subject testability was adequately attended to. Neither found a degree of improvement that altered the general conclusion on predictability.

There have been three efforts to identify test items having forecasting value. In the Fels series (Nelson & Richards, 1938, 1939; Richards & Nelson, 1939) the Gesell items on alertness and perception at 6, 12, and 18 months had promising correlations with later IQ and were superior to the whole scale in this. The Berkeley data (Hastings, 1952; Pinneau, 1953; summary by Bayley, 1955) have similarly been analyzed, yielding little of value before 18 months. After 2 years the prediction becomes surer. Tasks which are verbal or complex seem best. A third such study comes from Catalano and

McCarthy (1954), where measures of infant consonantal differentiation correlated about .40 with Binet later on, the magnitudes being little reduced when age of phonemic recording or of testing was partialed out.

The above paragraphs suggest that precious little developmental test work has come forth in many years of time. Ability testing, preoccupied with technical purity of verbal and nonverbal G, has chosen to eschew the imitative, psychomotor, and other functions which do not correlate well "with later IQ," omitting large areas of human competence which might have values in their own rights. Creative hours have been dedicated to projections of conative life.

For want of usable standardized instruments, the personnel in certain healing arts have had to resort to improvisation. Other productions are unpublished, being used in the specific hospital, clinic, or school. Both varieties show a range of quality and occasional inventiveness in specific items. Most are improvisations with existing materials, the Gesell items being the most favored. All share the shortcomings consequent to the lack of test-making art: (*a*) unclear directions for administration and scoring, (*b*) unspecified norms or none at all, (*c*) unawareness of how these shortcomings, together with "adaptation for local use," can affect the difficulty level of an item. Some instruments worthy of mention, in spite of the shortcomings, are those of Haeussermann (1958), Johnson, Zuck, and Wingate (1951), and Kogan (1957). The clinical psychologists working with such service personnel have tried to choose more carefully from existing instruments and have shown more awareness of the standardization problems (Gallagher, Benoit, & Boyd, 1956; Shontz, 1957). One must at this point remind the clinical psychologist that it was a physician who developed the Bender-Gestalt (Bender, 1938).

In spite of such demonstrated call for the services of psychological test makers, there has been little advance in either concept or practice in the ability testing of the young since the time of Binet. The elegant simplicity of the age-scale notion served the needs well for several decades. It is, however, no longer possible to defend narrow spectrum G testing. Voices have appealed for change (Bayley, 1955, 1956; Sarason & Gladwin, 1958; Thurstone, 1956). But the only energies exerted toward new scales which the present writers have knowledge of seem limited to those of students of Kirk and Osgood at Illinois (e.g., Sievers, 1955).

Potential Contributions of Factorial Study

Beyond the need for greater clinical breadth and prediction, the growth studies could very well use some instruments which provide whatever continuity is available from one age to another in the functions tested. Even more intriguing than continuity is the emergence and differentiation of those abilities which one calls "factors" when the child grows up.

The study of abilities at older levels has enjoyed considerable sophistication. A fairly stable structure of human competencies is taking form in replicated works (Fleishman, 1957; French, 1957; Guilford, 1956, 1957, 1958). Factorial study is not in its infancy. Its contributions no longer need to be in the form of scattered dramatic discoveries, but are systematic, stepwise contributions. Not all the expected factors of adulthood have yet been operationally described, nor have all the

TABLE 2

FACTORS REPORTED IN STUDIES PERFORMED AT PRESCHOOL AGES

Richards & Nelson (1939)	McNemar (1942) 14 Analyses for 2 Years & Up			Hofstaetter (1954)			Kelley (1928)	Thurstone & Thurstone (1953) (Primary Mental Abilities)[a]
6, 12, 18 months	2–2½ years	3–4 years	5–6 years	0–2 years	2–4 years	3 years up	3½–6½ years	5–7 years
	G[b]	G	G				maturity = heterogeneity, etc.	(intercorrelations .5 & up)
alertness							spatial no. 1 spatial no. 2	perceptual speed space
	identifying							verbal
motor								motor
			a numeration factor					quantitative
							auditory memory span[c]	
	motor or memory[d]			sensorimotor alertness	persistence	intelligence	memory	

[a] This array is that of the published test; the factor study was not published.
[b] Factors with apparent likeness are horizontally aligned.
[c] Labeled "verbal" by Kelley but tasks were two auditory span lists.
[d] Test items: form board, block tower, digit repetition.

questions one raises about them or the structure they fit been answered. Whether the many factors have market value or how they emerge as nature-nurture products, are questions which for the moment are beside the main issue. If factorial descriptions of man's abilities are available, one can make efficient measurement of whatever attributes one chooses or needs for a specific purpose. The need might be to know whether immediate memory span is "fixed" or trainable, what attributes will efficiently predict successful home leave from an institution, or the anticipation and prevention of difficulties in primary reading. The main point is that a general ability or G, if it exists, is not greatly predictive from one young age to another. Certainly it is an insufficient description for a psychologist to make of the competencies of a person. On the other hand, factorial study may enable one to discern and differentially measure, with efficiency, the particular congeries of traits which suit a specific purpose.

It should be of advantage to seek for continuities between the established factor structure for young adults and the rather bulky amount of information which is available from the observations and experimental study of the infant and young child.

Previous Factor Studies

Attention is first given to those factorial studies of abilities known to have been performed at preschool ages. Only five, all of them American, have been found. A perusal of British summaries of factor literature (e.g., Thomson, 1950; Vernon, 1950) did not reveal evidence of any other preschool analyses. The findings of these studies are presented (Table 2), not to claim that some group factors have been discovered, replicated, and accepted, which is hardly the case, but only to show that something beyond a general factor can be found. Some discussion of the findings of the previous studies seems warranted.

Richards and Nelson (1939) analyzed the interitem tetrachoric *r*'s of the Gesell items in the Fels data. Clearly emerging on the centroid extractions at 6, 12, and 18 months

TABLE 3

FINAL FACTOR VALUES FOR THE POPULATION OF 107 KINDERGARTEN CHILDREN AT CLOSE OF FORTY-EIGHTH SUCCESSIVE APPROXIMATION*

Tests Used	Maturity, Hetero-geneity, etc.	Verbal	Memory	Spatial No. 1	Spatial No. 2	Control of Mean-ingless Content?
1. Memory for Meaningful Forms	.79		.13			
2. Control of Meaningful Visual Memory Images	.71			.27		−.18
3. Memory of Meaningless Forms	.62		.18	.27	.50	
4. Control of Meaningless Visual Memory Images	.63				.15	.42
5. Memory for Verbal Material	.49	.61	.50			
6. Divided Forms Test	.52				.58	−.24
7. Knox Cube Test	.62		.36		.13	

* From Kelley (1928, p. 148), with slight modification.

were two factors labeled "alertness" and "motor." At 12 months there was a hint of a third factor.

McNemar (1942) reported on the standardization of the 1937 Stanford-Binet. Included were analyses which, unlike most factor efforts, were designed to disclose whether the item selection of the famous scale was carefully enough done to have avoided group factors. The unrotated centroid analyses demonstrated at most age levels the excellence of item selection from the viewpoint given: a first loading accounted for most of the common factor variance at nearly all age levels. Exceptions included 2–0 and 2–6, at which there were two apparently unimportant loadings of unclear meaning, evident in the identifying, memory span, and movement items.

Hofstaetter (1954) analyzed a matrix of interage test correlations. The results showed a "sensorimotor" alertness with best loadings in the first two years, "persistence" from two to four, and "intelligence" from three years on. Thus a sophisticated treatment confirms what is seen in the inspectional analysis of test content and the study of age-to-age intercorrelations.

The above-mentioned three studies were analyses of already existing data. Only Kelley's and the Thurstones' analyses sought for factors. Kelley (1928) was the pioneer and provided factor reports for seventh grade, third grade, and kindergarten children. With a highly rational but painstaking and difficult method no longer used, he reported on the upper two groups a pattern that we today would identify as verbal, numerical, spatial, etc., with loadings also on what he prefers to call "maturity-heterogeneity" rather than G. Only his kindergarten array is of interest here. The original table is substantially reproduced in Table 3.

The findings deserve some discussion. First, Kelley was one of the few in his day in resisting G, believing the British data for it to be due to insufficiently controlled variability in

sex, age, and background of subjects and to unrealistic treatment of error in residuals. Kelley labels his own first factor as maturity-heterogeneity (his age range at kindergarten was 3–6 to 6–3). Second, what he called "verbal" should be called "memory span," or "immediate memory," as the items required only serial recital after one hearing, not semantic interpretation. This "verbal" (or memory) is found for only one entry in the matrix.

Leaving aside the characteristics of the work in today's terms, the Kelley report is not contradicted by later data. There is replication to the extent the Thurstones, with group tests, explored down to age 5½. The Kelley study remains the best available model for anticipation of factors.

The second and remaining analysis-for-factors' sake at preschool ages was the work of the Thurstones leading to the PMA tests (Thurstone & Thurstone, 1953) which included a level for ages 5–7. According to the publisher,[2] the factorial analysis of the 5–7 age group has not been published; the norms for the published tests were established on a new group. The 5–7 PMA has this array: motor, perceptual-speed, verbal, spatial, and quantitative (replaced by N at higher levels). "Motor" is tested by requiring the drawing of pencil lines between dots and compares to "dynamic precision" at higher levels. It is not in the array of the PMAs for older subjects. Regarding the Q or quantitative, the Thurstones point out that N and R evolve from Q sometime after the kindergarten-primary level.

While the original factor study was not published, the Technical Supplement to the Manual provides some information on groups of pupils utilized for correlational studies. There is less evidence of factor clarity at 5–7 compared with the higher ages. Intercorrelations were .50 and up. With an age range of only 5–9 to 6–8, the prospect of reduction by partial r is small. Another table shows r's with Binet; only motor is below .50.

The PMAs cannot be considered replications of specific Kelley factors, but the two reports do reinforce the conviction that factors can be found. Note that Kelley had a first loading called maturity-heterogeneity and the Thurstones found considerable interfactor correlation. To label either result the consequence of G is as unwarranted at this time as to conclude the cause is differential testability or variance due to testing conditions. The McNemar report had little common factor variance beyond the first loading, but in this instance "purity" was built into the test material by preliminary steps. The general conclusion, therefore, is that factors will emerge once appropriate test materials are made available.

Problems in Conducting Factorial Studies with the Young

That only five studies have been located is not entirely due to lack of interest. The problems of testability at preschool ages are discouraging. It is not until fourth or fifth grade (age 9 or 10) that a typical group of children have enough reading ability and conformity to be tested easily in full class groups. In second and third grades (and some fourth) one must read directions aloud with the children and provide monitors; even so, only 30 minutes of adequately controlled testing are accomplished. One who watches the process might chal-

[2] Robert H. Bauernfeind, Personal communication, February 18, 1959.

lenge the word "adequately." At kindergarten it is the custom to do readiness evaluation in groups not exceeding four or five. School psychologists are not at all sure that even this situation gives control over testing circumstances, especially when other personnel than themselves do the administering. It is almost certain that individual differences in conformity, distractibility, broken pencils, and the like have contributed to variance and lie behind the larger test and factor intercorrelations found in the younger level when different levels have been compared. The skeptic should watch the PMA being administered even to a small, well-motivated kindergarten-primary group. Below kindergarten, of course, no useful testing other than individual can be accomplished, and the cost factor becomes significant.

Hypothesized Factors

Introduction

It is convenient to return briefly to the issue of prediction from an early to a later age. The discovery of factors at age 5 is of merit in itself, whether or not what is found has continuity with abilities at other levels. Relation with findings at higher age levels can be taken for granted. It is another matter with respect to extension downward from age 5 to infancy. There is little guarantee that any particular ability seen at age 5 is represented at 3 or earlier. In other words, the lack of prediction from baby tests, previously discussed, may not entirely be due to poor instrumentation; to some degree, qualitative difference exists between infant and child. The issue of whether abilities emerge via differentiation or via consolidation certainly enters. Bayley (1955, 1956) provides good discussion of this. But answers to these and related questions cannot be produced till the instrumentation capable of detecting what exists has been created.

A further introductory note is required. The word "ability" needs more than an implicit definition. It is here regarded as that functioning which does *not* include (a) the so-called vestigial reflexes such as the plantar; (b) the vegetative responses even if they can be volitionally controlled at times, such as sphincter activity; (c) random, apparently unaimed "emitted" movement of the skeletal sets, such as the arm and leg motions of the infant.

It is more difficult to say what *is* included, but tentatively, it is behavior change which is guided by current or previous sensorial input. For example, cessation of movement at a sound is regarded here as an ability. This breadth of concept is wanted in order to include within the scope of human "adaptive ability" any movement, change in movement, cessation of movement, readiness for movement, etc., when such conduct is performed in direct or delayed consequence of experiencing with the senses. Of necessity, such movements and their attributes must include such nonintellectual dimensions as strength, simple reaction time, and the like.

For clarity's sake, the above discussion and the test items to be mentioned might be thought of in the traditional expression: S-O-R.

As usual, the term S represents the stimulus situation or any portion of it we choose to center upon, current or past. As usual, R represents a response, a muscular activity or change in same, of any recordable or observable sort. Discussion of O is postponed for the moment.

In the use of a test of perceptual

functioning, it is hoped that the obtained individual differences were due to perception, not to O or R. A response must be made by the subject, of course. But the examination process provides that there is simplicity and ease in the means of response. Hence in a perceptual task a subject signifies his recognition or discrimination by a gesture, a lever depression, a word, or a pencil mark. On the other hand, if one seeks individual differences in response speed, strength, dexterity, sentence length, imaginativeness, or other quality, then the input factors should bring no differential difficulty to the subjects.

With respect to the O term above, the present paper needs to presume no special theoretical position. The intermediation between input and response can stand for "mind" or physiology. The expression "thinking factors" is utilized below, and refers without commitment to whatever intermediates. Test items which have their difficulty in "mental" or thinking intermediators should put no burden upon the perceptual or response operations of the examinees. Hence, one uses a time or space gap between S and R; one requires the subject to draw a conclusion on similarity or difference, to extrapolate, to find a rule, to make a judgment, to find new uses or combinations.

The hypothesized factors now presented are grouped, to some extent, on the grounds just presented.

Table 4 presents domains and hypothesized factors within them, for ages 4–6. It also shows estimated ages of emergence of the proposed abilities as identifiable factors. The table also lists sources in the literature which implicitly or explicitly support certain of the differentiations assumed in the array.

Three domains ought to be fairly certain of existence: psychomotor, perceptual, and psycholinguistic. The seven which are specified result from the separation of auditory from visual perception, the separation of gross-body psychomotor from hand-eye, the separation of receptive from expressive psycholinguistics, and the differentiation of mental or thinking from all of these.

It is not claimed that all the hypothesized factors can be found. No immediate claims about orthogonality can be made. Until the first few simple matrices are analyzed, one can make only guesses.

The instruments listed in Table 1 (including the school readiness tests and the differentiated batteries) have abundant raw material for the construction of tentative scales having some promise of purity. The research literature (e.g., Gewirtz, 1948a, 1948b) provides further ideas for items seeming to represent relatively narrow spectrums of ability like those of "reference" factor tests. Hence, some item sources pertinent to the speculated abilities have been entered into Table 4.

The Motor Domains

Two general groups of motor factors are presumed: (a) the whole-body, in which gravity is defied in postural and locomotor achievements and (b) the hand-eye or manipulational. The developmental psychologist will quickly note that the motor items in baby scales, in the Bayley-California (1935) and the Oseretsky (Sloan, 1955) can be easily subdivided into whole-body and hand-eye items. Gesell did this a generation ago in separately listing motor and adap-

TABLE 4

FACTORS HYPOTHESIZED IN SEVEN DOMAINS WITH RELATED INFORMATION

Early Appearance	Hypothesized Factors, CA 4–6 Years	Age of Emergence of Factors	Tests & Items	Test & Item Sources	Bibliography pertaining to Domain
DOMAIN 1: Psychomotor, whole-body *1 month:* Earliest antigravity, chin-up; momentary head control *5 months:* Rolls over *7 months:* Erect sitting *8 months:* Crawling progress *12 months:* Standing alone *12–15 months:* Walking *24 months:* Up and down stairs	1.1 Postural balance	1½ years	Heath rail walk; toe-balance; chalk walk	Bayley (1935) Glanville & Kreezer (1937) Heath (1953) Hempel & Fleishman (1955) McCloy & Young (1954) Metheny (1941) Sloan (1955)	Bayley (1935) Fleishman (1953, 1957) Guilford (1958) Jones (1949) Jones & Seashore (1944) McCloy & Young (1954)
	1.2 Dynamic balance	1½ years	Hop-skip; one-foot hop; jump-balance		
	1.3 Impulsion	2½ years	Standing broad jump & high jump; football kick		
	1.4 Coordination	2½ years	Hurdle jump; cable jump; rotary pursuit		
	1.5 Flexibility	3 years	Back-down-wall; toe-touch; amplitude of joint movement		
	1.6 Strength	3 years	Dyanometer; abduction of leg; push-ups		
DOMAIN 2: Psychomotor, hand-eye *3 months:* Retentive grasp *4 months:* Grasps and secures *7 months:* Drinks from cup *10 months:* Pulls toy by string *30 months:* Buttoning	2.1 Static precision	3 years	Steadiness-aim; track tracing	Bayley (1935) Glanville & Kreezer (1937) Heath (1953) Hempel & Fleishman (1955) McCloy & Young (1954) Metheny (1941) Sloan (1955)	Bayley (1935) Fleishman (1953, 1957) Guilford (1958) Jones (1949) Jones & Seashore (1944) McCloy & Young (1954) Seashore (1951)
	2.2 Dynamic precision	3 years	Rotary pursuit; pursuit-aiming; circle dotting; hole punching		
	2.3 Reaction time	3 years	Auditory & visual RT		
	2.4 Dexterity	3 years	Tweezer tasks; cube stacking; nut stacking		
	2.5 Speed	4 years	Dot tapping; finger tapping; articulation speed (also see 6.1)		

TABLE 4—(*Continued*)

Early Appearance	Hypothesized Factors, CA 4–6 Years	Age of Emergence of Factors	Tests & Items	Test & Item Sources	Bibliography pertaining to Domain
DOMAIN 3: Visual perception *1 month:* Awareness—pursuit movements *5 months:* Recognition of parents, differentiation from strangers *6 months:* Other percepts	3.1 Perceptual speed	18 months	Thurstone PMA tests; readiness items; letter and word discrimination; Leiter items; Columbia items; (Burgemeister)	Burgemeister et al. (1954) Lee & Clark (1951) Leiter (1948) Monroe (1935) Templin (1957)	French (1957) Gollin (1956) Guilford (1956, 1958) Munn & Steining (1931) Thurstone (1948) Thurstone (1956) Thurstone & Thurstone (1953)
	3.2 Space	36 months	Thurstone PMA tests; space items in readiness tests	Terman & Merrill (1937) Thurstone & Thurstone (1953)	
DOMAIN 4: Auditory perception *1 month:* Awareness *5 months:* Auditory localization *6 months:* Identification of mother, bottle sounds, voices *9 months:* Word sound discrimination; differential reaction to sounds, voices, tone of voices	4.1 Auditory discrimination	1½ years	Identification of common sounds; phoneme discrimination; discrimination learning	Lee & Clark (1951) Monroe (1935) Sievers (1955) Templin (1957)	Osgood & Sebeok (1954) Sievers (1955) Templin (1957)
	4.2 Auditory localization	1½ years	Indicating sound sources	Terman & Merrill (1937)	
DOMAIN 5: Receptive psycholinguistics *1 month:* Quieted by voice *3 months:* Responsive vocalization to voice *8 months:* Responds to name and "no" *18 months:* Identifies pictures by name	5.1 Auding	1 year	Ammons vocabulary; Stanford-Binet identifying items; use identification items; directions tests; sentences with prepositional instructions	Ammons & Ammons (1948) Gollin (1956) Hildreth & Griffiths (1949) Monroe (1935)	Brown (1954) Caffrey (1955) Eisenson (1954) Gewirtz (1948a, 1948b) McCarthy (1954) Osgood & Sebeok (1954) Sassenrath & Holmes (1956)
	5.2 Verbal comprehension	5 years	"Preliterate" items; recognition of own name; address; meaning of common signs, symbols, numbers at "readiness" level; cartoon meaning; Thurstone street-gestalt; Sievers object–picture differential vocabulary	Sievers (1955) Stutsman (1931) Templin (1957) Terman & Merrill (1937) Thurstone (1948)	

TABLE 4—(*Continued*)

Early Appearance	Hypothesized Factors, CA 4–6 Years	Age of Emergence of Factors	Tests & Items	Test & Item Sources	Bibliography pertaining to Domain
DOMAIN 6: Expressive psycholinguistics *1 month:* Noncrying vocalization *6 months:* Babbling *12 months:* First spoken word	6.1 Articulation	1½ years	Speed of repetitive pronunciation different phonemes	Gewirtz (1948a, 1948b) Monroe (1935) Sievers (1955) Stutsman (1931) Templin (1957) Terman & Merrill (1937)	Eisenson (1954) Gewirtz (1948a, 1948b) McCarthy (1954) Osgood & Sebeok (1954) Thurstone (1948)
	6.2 Semantic fluency	2 years	Naming or use vocabulary, thing names, food names, child names, etc.; picture description; Templin sentence complexity; action-agent; uses of objects; Sievers picture vocabulary; Binet picture vocabulary; how many? items		
	6.3 Symbolic fluency	3 years	Sievers nonsense imitation; alliteration; rhyming; word-sound games		
DOMAIN 7: Mental (memory & thinking) *8–9 months:* Piaget's "fourth stage" *10 months:* Imitation *15 months:* Discrimination learning (Munn & Steining, 1931) *30 months:* Rotated form board	7.1 Memory span	2½ years	Digit span; word span; sentence memory; Sievers nonsense grammatical mimicry; Knox cubes, Binet picture memory and design memory	Gollin (1956) Keller (1954) Monroe (1935) Sievers (1955) Stutsman (1931) Terman & Merrill (1937)	Bayley (1955, 1956) Garrett (1946) Gollin (1956) Keller (1954) McCarthy (1957) Maurer (1946) Piaget (1952) Thurstone (1948) Thurstone (1956) Werner (1948)
	7.2 Abstracting	3 years	Verbal analogies; sorting; transfer tasks		
	7.3 Reproduction of visual models	3 years	Kohs blocks; Rutgers drawing; bead chain; Thurstone pattern completion and pattern copying		

tive, while Griffiths (1954) utilized similar rubrics. In a more sophisticated way, Guilford (1958) provides a structure which distinguishes "gross" psychomotor factors from those of specified parts of the body.

Taking useful reaction to gravity as the first expression of psychomotor ability, then the earliest response occurs in the first month. The normal baby shows momentary maintenance of head posture when held erect, and can lift his face from prone and turn it so as to breathe. The onward development includes postural and locomotor achievements against gravity. There being no factor studies at preschool ages to give leads, the observations of Gesell and others must remain the best sources of the guesses. Whole-body motor differentiated from hand-eye within the first year, with low congruence believed to occur before two years. Among whole-body activities, the antigravity and the locomotor are different by "face" appearance, and clearly separate in adulthood (Guilford, 1958). This distinction appears in the words "static" and "dynamic" in Table 4. Within each of the whole-body and hand-eye domains, differentiation should be easily observable before school age. In fact, perusal of Table 4 shows that greater differentiation is predicted in the psychomotor than in other domains. This particularization is suggested by the low intercorrelations of various motor tests (Jones & Seashore, 1944). Adult motor abilities are surprisingly specific (Seashore, 1951) and differentially trainable (Fleishman, 1957).

There are likewise no factorial studies for later preschool age to utilize for prediction. There are normative observations of motor progress till about first grade, and data from physical education, fitness testing, and the Oseretsky scale, thereafter. What is known factorially comes from the physical education literature (e.g., McCloy & Young, 1954) and from "human engineering" (e.g., Fleishman, 1953, 1957). The subjects of the former are secondary school or college learners and the domain is usually whole-body performance; the latter typically has adult subjects and a hand-eye domain. A "psychomotor structure" presented by Guilford (1958) puts the factor work into one perspective and is the basis for the motor domains proposed here.

There are age and sex differences in times of reaching maxima. The development in girls in vigorous athletic performance goes into a plateau or declines at 12 or 13 years, while that of boys continues to gain till 18 or 20 years (Espenshade, 1940; Jones, 1949). Measures of balance and coordination plateau for both sexes in adolescence (Jones & Seashore, 1944) while muscular strength grows apace in both sexes but more so in boys, for more years (Jones, 1949). Hence, while it is too early to anticipate the ages of the various factor appearances, strength should certainly be one of the first to emerge, as a factor separate from those mentioned and within both whole- and part-body activities.

To sum up on motor factors, whole- vs. part-body should be evident within 2 years; a nearly full display like the adult structure of Guilford is expected by 5 years of age.

Visual Perception

The utilization of exteroception and somesthesis is observed within the first weeks of life. One may see primitive awareness or discrimination at 1 month or earlier, in a change in conduct at a change in the stimulus

field (visual pursuit, cessation of movement at a sound, etc.). At 3 or 4 months the primitive awareness has become recognitive, perhaps only at a conditioning level, by the criterion of anticipatory behavior at sight of bottle or mother. The word "percept" is deserved at 6 months if not earlier in that the infant seems to be able to carry a recognition past the immediate time or space. Discriminating, recognizing, and percept building must precede, in all logic, the use of receptive input when that use is more than reflexive. The order of items in the Gesell scale, for example, shows recognitive behavior before adaptive positive response.

A distinction between perception and spatial ability should be evident at 3 years, according to Thurstone (1948). Kelley's study identified two visual factors, one probably spatial, in subjects whose ages reached down to 3½. The Merrill-Palmer (Stutsman, 1931) and the Stanford-Binet (Terman & Merrill, 1937) have items which appear to be both of the speedy differentiating kind and of the visualization sort at several levels. In a study relating reading to various Thurstone perceptual tests, Goins (1958) identified two factors. The first appeared mainly in the timed tests while the second might be called spatial, as it appeared on untimed tasks of pattern completion, copying, reversals, etc. Of pedagogical significance is that the first grade reading test loaded on the two factors .346 and .609, respectively. If these loadings surprise the noneducationist, it should be recognized that primary reading is largely "saying words out loud" with little or no V at the outset. At older ages other studies (except a few British ones; see Vernon, 1950) established that space is separable from P, and the complexity becomes considerable in adulthood (French, 1957; Guilford, 1956).

In summary, therefore, it is proposed that visual perception will be seen as noncongruent with motor and other abilities at 6 months and that by 3 years will differentiate into separably measurable if not orthogonal perceptual vs. space factors.

Auditory Perception

It is necessary to point out that the distinction made here between visual and auditory perception is on theoretical grounds. The classic Thurstone (1938) work on perception was exclusively on visual. The subject of channel differences intrigues the communication people (Osgood & Sebeok, 1954) and was suggested by Thurstone himself (1948).

Evidence for more than reflexive use of auditory input can be detected as early as evidence for similar use of visual input. That is, within the first month, a normal baby is quieted by sound or otherwise changes his activity. However, it is less clear how auditory skills proliferate for the necessary testing confuses auditory discriminatory skills with word semantics. While a child who selects "bear" separately from "pear" in a "spondee test" clearly shows ability to distinguish the two initial sounds, one cannot tell how early this perceptual distinction occurs by simple testing before a good "auding" vocabulary has been acquired. It is necessary to use differential conditioning or discrimination learning. It is therefore necessary, in practical testing, to ignore auditory figural discrimination vs. auditory semantic factors at their earliest and to limit present predictions of high refinement of factors to about second grade testability.

But if the resources for extensive investigation become available, it is predicted that such factors analogous to the visual can be detected by 2 years and before: space orientation via sound, differential reaction to word-sound phonemes apart from meaning, pitch discrimination, and many others.

Linguistic

Receptive. The distinction between receptive and expressive language was known in the literature of aphasia before Thurstone distinguished W from V (cf. French's review, 1951). The distinction is demanded in communications theory (e.g., Osgood & Sebeok, 1954). In addition, channel differences are also suggested by these theorists as well as by the auding literature from educational psychology (Brown & Caffrey, 1952; Caffrey, 1955; Sassenrath & Holmes, 1956) and experience with slow learners (Durrell & Sullivan, 1958; Kirk & McCarthy, 1950). Thus, one must not yet assume that an auding factor in prereading ages is congruent with later V developed with printed tests.

Receptive factors are hypothesized as shown in Table 4. The auditory channel includes a factor of speech-sound perception. Difficulties in testing for it were mentioned just above. One might anticipate it at those levels at which it must be presumed present—before 1 year of age when a child shows evidence of differential understanding of words before he can use them expressively. The hypothesized visual comprehension factor would be only primitive as yet. Its origins may be detectable at ages 5 and up by items from readiness tests in which meaning is obtained, for example, in specifying the significance of a traffic signal. Note that both auditory and visual perceptive factors are located in both linguistic and the perceptual domains.

Expressive. The array here also distinguishes the vocal from the graphic-gestural modes of expression. Since the age of interest here is preschool, the written language could at best be rudimentary; at kindergarten the ability to hold a pencil, use it, and print one's name is generally expected, little more. Psychomotor factors which are listed under appropriate heading are repeated here.

Thinking

Referring back to the S-O-R discussion, "thinking" refers to that which is intermediate between discrimination of the stimulus matrix and the making of the response. The intermediation may be memory or visualization, rule-finding or abstracting. In any case, a test item needs to have its difficulty in O and not in stimulus discrimination or response refinement, and certainly not in word knowledge as such. For children of 5 or 6, the verbal problem is best avoided by simple omission. The choice of item content is limited largely to elemental figural materials among which discriminations are known to be made in the first year of life.

Hypothesized are memory span, abstracting, and reproduction of figural models. The last-mentioned is probably complex; in adult work the Kohs and the pattern copying tasks are loaded on space but also on reasoning, perceptual, verbal, and specific factors (French, 1951). Obviously it is going to cause some problems to separate space from thinking and perhaps this task cannot be accomplished so long as figural materials are used.

Memory span of course is easily obtained by both verbal and gestural means, provided the contents are "easy" to begin with. Digit span for the normally cultured child is satisfactory. The Knox cubes provide another span-measuring process of reasonably apparent purity.

Further speculation about the thinking factors is probably premature. There is insufficient experience at preschool ages with analogs of, for example, what Guilford would call productive thinking. Existing Binet and WISC items are of three main sorts: immediate memory span; perceptual discriminations and identifications; and information "bits" brought forth by items of the sort, What should you do when, Which is prettier, How many, etc.

It is not till the age-levels of 6 and 7 that the Stanford-Binet uses productive thinking items. At these age-levels one finds verbal analogies, picture absurdities, and similarities and differences. Defining thinking as the processing of information to produce a conclusion, then thinking is so introduced. Such items are concerned with semantic content. Symbolic content also is rare at younger levels. It was already pointed out that one competent team found no "number," a symbolic ability, in the age bracket 5 to 7 years (Thurstone & Thurstone, 1953). The making of rhymes, which would be another ability in symbolic content is called for in the Stanford-Binet at the 9-year level.

The WISC (Wechsler, 1949) is another scale in common use which, like the Binet, essentially lacks thinking items, as above described, below the mental-age equivalent of about 7. The first items in the Arithmetic subtest require only counting. But the WISC has some verbal analogies with mental-age values at around 5 and 6 years.

Items calling for "divergent" productive thinking are almost nonexistent in mental tests. The WISC has none. The Binet (Form L) calls for word-saying at age 10. Monroe's (1935) old readiness test calls for naming of objects in specified classes (e.g., animals).

Hence the preschool age-levels of popular batteries are barren of thinking items; the speculator has few reference points on which to develop a system. The many observations of Piaget should not be ignored, nor those of American investigators on problem solving in the young (a recent example, Braine, 1959). But the aggregate of items and ideas from these scattered sources is not large. A considerable amount of inventive elaboration is required if even a modest 6-year-old model of adult-level structure is wanted.

Summary

This paper has reviewed the situation of ability testing at the infant and preschool ages. The available instruments have, on the one hand, failed to predict "later IQ," for whatever value later IQ has as a criterion. On the other hand, the spectrum of abilities tested, which if anything should broaden as the child grows, is caused to narrow down by the use of technical criteria of age-progression and internal consistency in item selection. The previous factorial studies at preschool ages, of which only one (Kelley's) may be regarded as of quality, give reason to believe that carefully prepared testing could demonstrate that more than a G is present. Finally, a series of factors is hypothesized, representing seven domains of whole-body and hand-eye

psychomotor, visual and auditory perception, receptive and expressive psycholinguistics, and thinking or mental. The hypotheses are made for ages 4–6 years, with some speculations as to age of earliest emergence of the factor. Illustrative item-types and supporting literature are given.

REFERENCES

AMMONS, R. B., & AMMONS, H. S. *Full-Range Picture Vocabulary Test.* Missoula, Montana: Psychological Test Specialists, 1948.

ARTHUR, GRACE. *A point scale of performance tests.* (Rev. Form II) New York: Psychological Corp., 1947.

BAYLEY, NANCY. The development of motor abilities during the first three years. *Monogr. Soc. Res. Child Develpm.*, 1935, No. 1.

BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist*, 1955, **10**, 805–818.

BAYLEY, NANCY. A new look at the curve of intelligence. In, *Proceedings of 1956 Invitational Conference on Testing Problems.* Princeton: Educational Testing Service, 1956.

BENDER, LAURETTA. A visual motor gestalt test and its clinical use. *Amer. Orthopsychiat. Ass. res. Monogr.*, 1938, No. 3.

BRAINE, M. D. S. The ontogeny of certain logical operations: Piaget's formulation examined by nonverbal methods. *Psychol. Monogr.*, 1959, **73**(5, Whole No. 475).

BROWN, D. Auding as the primary language ability. Unpublished doctoral dissertation, Stanford University, 1954.

BROWN, D., & CAFFREY, J. G. *California Auding Test.* (Rev. ed.) Redwood City: Council on Auding Research, 1952.

BURGEMEISTER, BESSIE B., BLUM, LUCILLE H., & LORGE, I. *Columbia Mental Maturity Scale.* Yonkers, New York: World Book, 1954.

CAFFREY, J. Auding. *Rev. educ. Res.*, 1955, **25**, 121–138.

CATALANO, F. L., & MCCARTHY, DOROTHEA. Infant speech as a possible predictor of later intelligence. *J. Psychol.*, 1954, **38**, 203–209.

CATTELL, P. *The measurement of intelligence of infants and young children.* New York: Psychological Corp., 1940.

CAVANAUGH, MAXINE C., COHEN, I., DUNPHY, D., RINGWALL, E. A., & GOLDBERG, I. D. Prediction from the Cattell Infant Scale. *J. consult. Psychol.*, 1957, **21**, 33–37.

DOLL, E. *Vineland Social Maturity Scale.* New York: Psychological Corp., 1953.

DURRELL, D. D., & SULLIVAN, HELEN B. Newsnote. USDHEW Office of Education, Cooperative Research Program, 1958, Sept. (Reprinted from *Sch. Life*)

EISENSON, J. *Examining for aphasia.* (Rev. ed.) New York: Psychological Corp., 1954.

ESCALONA, SIBYLLE. The use of infant tests for predictive purposes. *Bull. Menninger Clin.*, 1950, **14**, 117–128.

ESPENSCHADE, ANNA S. Motor performance in adolescence. *Monogr. Soc. Res. Child Develpm.*, 1940, **5**, No. 1.

FLEISHMAN, E. A. Testing for psychomotor abilities by means of apparatus tests. *Psychol. Bull.*, 1953, **50**, 241–262.

FLEISHMAN, E. A. A comparative study of aptitude patterns in unskilled and skilled psychomotor performances. *J. appl. Psychol.*, 1957, **41**, 263–272.

FRENCH, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometr. Monogr.*, 1951, No. 5.

FRENCH, J. W. The factorial invariance of pure-factor tests. *J. educ. Psychol.*, 1957, **48**, 93–109.

GALLAGHER, J. J. Clinical judgment and the Cattell Infant Intelligence Scale. *J. consult. Psychol.*, 1953, **17**, 303–305.

GALLAGHER, J. J., BENOIT, E. P., & BOYD, H. F. Measures of intelligence in brain damaged children. *J. clin. Psychol.*, 1956, **12**, 69–72.

GARRETT, H. E., A developmental theory of intelligence. *Amer. Psychologist*, 1946, **1**, 372–378.

GESELL, A. The ontogenesis of infant behavior. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1954. Pp. 335–373.

GESELL, A., & AMATRUDA, CATHERINE S. *Developmental diagnosis.* (2nd ed.) New York: Hoeber, 1947.

GESELL, A., CASTNER, M., THOMPSON, H., & AMATRUDA, CATHERINE S. *Biographies of child development: The mental growth careers of eighty-four infants and children.* New York: Hoeber, 1939.

GEWIRTZ, J. L. Studies in word fluency: I. Its relation to vocabulary and mental age in young children. *J. genet. Psychol.*, 1948, **72**, 165–176. (a)

GEWIRTZ, J. L. Studies in word fluency: II. Its relation to eleven items of child be-

havior. *J. genet. Psychol.*, 1948, **72**, 177–184. (b).

GLANVILLE, D., & KREEZER, G. Deficiences in amplitude of joint movement associated with mental deficiency. *Child Develpm.*, 1937, **8**, 129–138.

GOINS, JEAN T. Visual perceptual abilities and early reading progress. *Suppl. educ. Monogr.*, 1958, **87**, Feb.

GOLLIN, E. S. Some research problems for developmental psychology. *Child Develpm.*, 1956, **27**, 223–235.

GRIFFITHS, RUTH. *The abilities of babies.* New York: McGraw-Hill, 1954.

GUILFORD, J. P. The structure of intellect. *Psychol. Bull.*, 1956, **53**, 267–293.

GUILFORD, J. P. A revised structure of intellect. *Rep. psychol. Lab.*, 1957, No. 19.

GUILFORD, J. P. A system of the psychomotor abilities. *Amer. J. Psychol.*, 1958, **71**, 164–174.

HAEUSSERMANN, ELSE. *Developmental potential of preschool children.* New York: Grune & Stratton, 1958.

HASTINGS, H. J. The predictive value of individual items in preschool intelligence tests. Unpublished doctoral thesis, University of California, 1952.

HEATH, S. R., JR. The relation of rail-walking and other motor performance of mental defectives to mental age and etiologic type. *Train. sch. Bull.*, 1953, **50**, 119–127.

HEMPEL, W. E., JR., & FLEISHMAN, E. A. A factor analysis of physical proficiency and manipulative skills. *J. appl. Psychol.*, 1955, **39**, 12–16.

HILDRETH, GERTRUDE H., & GRIFFITHS, NELLIE L. *Metropolitan Readiness Tests.* Yonkers, New York: World Book, 1949.

HOFSTAETTER, P. R. The changing composition of "intelligence": A study in T-technique. *J. genet. Psychol.*, 1954, **85**, 159–164

JOHNSON, M. K., ZUCK, F. N., & WINGATE, KATHLEEN. The motor age test: Measurement of motor handicaps in children with neuromuscular disorders such as cerebral palsy. *J. bone jt. Surg.*, 1951, **33**, 688–707.

JONES, H. E. Motor performance and growth: A developmental study of static dynanometric strength. *U. Calif. Publ. child Develpm.*, 1949, No. 1.

JONES, H. E., & SEASHORE, R. H. The development of fine motor and mechanical abilities. In N. B. Henry (Ed.), *Yearb. Nat. Soc. Stud. Educ.*, 1944, 43, Part I.

KELLER, J. E. The investigation of a specific cognitive deficiency. *Amer. J. ment. Defic.*, 1954, **58**, 560–565.

KELLEY, T. L. *Crossroads in the mind of man.* Stanford: Stanford Univer. Press, 1928.

KIRK, S. A., & MCCARTHY, J. J. A study of the language process of preschool cerebral palsied children. Progress Report to the United Cerebral Palsy Foundation, 1950.

KOGAN, K. L. A method of assessing capacity in preschool cerebral palsied children. *J. clin. Psychol.*, 1957, **13**, 54–56.

LEE, J. M., & CLARK, W. W. *Reading Readiness Test.* Los Angeles: California Test Bureau, 1951.

LEITER, R. G. *International Performance Scale.* Washington D. C.: Psychological Service Center Press, 1948.

MCCARTHY, DOROTHEA. Language development in children. (2nd ed.) In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1954. Pp. 492–630.

MCCARTHY, J. J. Difference in the language abilities of young cerebral palsied children. Unpublished doctoral dissertation, University of Illinois, 1957.

MCCLOY, C. H., & YOUNG, N. D. *Tests and measurements in health and physical education.* (3rd ed.) New York: Appleton-Century-Crofts, 1954.

MCNEMAR, Q. *The revision of the Stanford-Binet scale.* Boston: Houghton Mifflin, 1942.

MAURER, KATHARINE M. *Intellectual status at maturity as a criterion for selecting items in preschool tests.* Minneapolis: Univer. Minnesota Press, 1946.

METHENY, ELEANOR. The present status of strength testing for children of elementary school and preschool age. *Res. Quart.*, 1941, **12**, 114–130.

MONROE, M. *Reading aptitude tests.* Boston: Houghton Mifflin, 1935.

MUNN, N. L., & STEINING, B. R. The relative efficacy of form and background in a child's discrimination of visual patterns. *J. genet. Psychol.*, 1931, **39**, 73–90.

NEILON, PATRICIA. Shirley's babies after fifteen years: A personality study. *J. genet. Psychol.*, 1948, **73**, 175–186.

NELSON, V. L., & RICHARDS, T. W. Studies in mental development: I. Performance on Gesell items at six months and its predictive value for performance on mental tests at two and three years. *J. genet. Psychol.*, 1938, **52**, 303–325.

NELSON, V. L., & RICHARDS, T. W. Studies in mental development: III. Performance of twelve months old children on the Gesell schedule and its predictive value for mental status at two and three years. *J. genet. Psychol.*, 1939, **54**, 181–191.

OSGOOD, C. E., & SEBEOK, T. A. (Eds.) Psycholinguistics: A survey of theory and re-

search problems. *J. abnorm. soc. Psychol., Suppl.*, 1954, 49, 203.

PIAGET, JEAN. *The origins of intelligence in children.* (Trans. by Margaret Cook) New York: International Univer. Press, 1952.

PINNEAU, S. R. Changes in performance on intelligence tests from one month to eighteen years. Paper read at Western Psychol. Ass., Seattle, 1953.

RICHARDS, T. W., & NELSON, V. L. Abilities of infants during the first eighteen months. *J. genet. Psychol.*, 1939, 55, 299–318.

SARASON, S. B., & GLADWIN, T. Psychological and cultural problems in mental subnormality: A review of research. *Genet. psychol. Monogr.*, 1958, 57, 3–290.

SASSENRATH, J. M. & HOLMES, J. A. Auding and psycho-educational variables. *Calif. J. educ. Res.*, 1956, 7, 99–104.

SEASHORE, R. H. Work and motor performance. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1341–1362.

SHIRLEY, MARY M. The first two years: A study of twenty-five babies. Vol. II. Intellectual development. *Child welf. Monogr.*, 1933, No. 7.

SHONTZ, F. C. Evaluation of intellectual potential in hemiplegic individuals. *J. clin. Psychol.*, 1957, 13, 267–269.

SHOTWELL, ANNA M., DINGMAN, H. F., & TARJAN, G. *Pacific State Hospital Numbers Concept Test.* Sacramento: Calif. State Dept. Ment. Hygiene, 1957.

SIEVERS, DOROTHY. Development and standardization of a test of psycholinguistic growth in preschool children. Unpublished doctoral dissertation, University of Illinois, 1955.

SLOAN, W. The Lincoln-Oseretsky Motor Development Scale. *Genet. psychol. Mongr.*, 1955, 51, 183–252.

STUTSMAN, R. *Mental measurement of preschool children.* New York: World Book, 1931.

TEMPLIN, MILDRED C. Certain language skills in children. *Child welf. Monogr.*, 1957, No. 26.

TERMAN, L. M., & MERRILL, M. E. *Measuring intelligence: A guide to the administration of the new revised Stanford-Binet Tests of Intelligence.* Boston: Houghton Mifflin, 1937.

THOMSON, G. *The factorial analysis of human ability.* Boston: Houghton Mifflin, 1950.

THURSTONE, L. L. The perceptual factor. *Psychometrika*, 1938, 3, 1–17.

THURSTONE, L. L. Psychological implications of factor analysis. *Amer. Psychologist*, 1948, 3, 402–408.

THURSTONE, L. L., & THURSTONE, THELMA G. *SRA Primary Abilities, for ages 5–7.* Chicago: Science Research Associates, 1953.

THURSTONE, THELMA G. Changing concepts of intelligence: Implications for test construction. *Proceedings of 1956 Invitational Conference on Testing Problems.* Princeton: Educational Testing Service, 1956, 26–37.

VALENTINE, C. W. *Intelligence tests for children.* (4th ed.) London: Methuen, 1950.

VERNON, P. E. *The structure of human abilities.* New York: Wiley, 1950.

WECHSLER, D. *Manual for the Wechsler Intelligence Scale for Children.* New York: Psychological Corp., 1949.

WERNER, H. *Comparative psychology of mental development.* (Rev. ed.) Chicago: Follett, 1948.

WITTENBORN, J. R. A study of adoptive children: II. The predictive validity of the Yale Developmental Examination of Infant Behavior. *Psychol. Monogr.*, 1956, 70(2, Whole No. 409).

(Received August 5, 1959)

GEORGE BANTA COMPANY, INC., [illegible]